

Vowel–Zero Alternations
in West Slavic Prepositions

DEPARTMENT OF LANGUAGES AND LITERATURES
GOTHENBURG SLAVIC STUDIES 4
2014

Vowel–Zero Alternations in West Slavic Prepositions

A Corpus Based Investigation
of Polish, Slovak and Czech

Morgan Nilsson



UNIVERSITY OF GOTHENBURG

Doctoral dissertation in Slavic languages,
University of Gothenburg, 2014-11-08.
Draft edition for public defence.

Gothenburg Slavic Studies No 4.

Distribution:

Department of Languages and Literatures
University of Gothenburg
P.O.Box 200
405 30 Gothenburg
Sweden

© Morgan Nilsson, 2014

ISSN 2000-1134

ISBN 978-91-981198-3-1

e-published:

<http://hdl.handle.net/2077/36304>

Cover layout: Thomas Ekholm.

Printed in Sweden by

Reprocentralen, Humanisten, Göteborgs universitet, 2014.

Abstract

Ph.D. dissertation at University of Gothenburg, Sweden, 2014.

Title: Vowel–Zero Alternations in West Slavic Prepositions. A Corpus Based Investigation of Polish, Slovak and Czech.

Author: Morgan Nilsson.

Language: English, with a Swedish summary.

Department: Department of Languages and Literatures, University of Gothenburg, P.O.Box 200, 405 30 Gothenburg, Sweden.

Key words: Polish, Slovak, Czech, vocalisation, prepositions, vowel-zero alternations, corpus, phonology.

ISSN 2000-1134

ISBN 978-91-981198-3-1

The aim of this thesis is to investigate the scope of the vowel–zero alternations in prepositions in the three major West Slavic languages, i.e. Polish, Slovak and Czech, and to formulate, within the Government Phonology framework, the general phonological principles responsible for the majority of these alternations.

Many prepositions in the Slavic languages have two competing forms, such as the Czech *v* / *ve* *in*. It has long been a recognised problem that the conditions regulating the use of these two forms cannot easily be formulated, a difficulty which has led to complications in both prescriptive language recommendations intended for native speakers and in the teaching of these languages to foreign learners.

The present investigation is based on extensive corpus data which have been studied primarily with quantitative methods, but also to some extent qualitatively. Data and previous research have also been derived from the literature on the subject. The contexts of the occurrences of the vocalised form of the prepositions have then been defined as precisely as possible, and the phonological principles responsible for the phonologically driven vocalisations are discussed and formulated.

A certain proportion of the alternations have also been found to occur on lexical, and not phonological grounds.

The findings of this thesis make it possible to present both language learners and native speakers with new, more precise and detailed information concerning the actual usage of the vocalised forms of prepositions. The thesis also provides some interesting language-specific observations on vowel–zero alternations in relation to the comparative and typological research on the subject.

Acknowledgements

My work on this investigation has been slowly growing since 1995, when I made my first attempts at scanning texts for a small corpus, and a few years later I began to collect texts from the Internet. When I started, the West Slavic national corpora were still rather small and not easily accessible, and I therefore needed to collect a substantial amount of data on my own and to find a good tool to extract the relevant data from these corpora. I was told about Word Smith at the Institute of the Czech National Corpus and was later able to encourage its creator, Mike Scott, to elaborate certain sorting components of the program in order to better fit my needs and above all to handle Slavic languages in a more precise way. For his willingness to meet my needs I am greatly indebted to him.

Many of the ideas presented in the book first came to my mind during my stay at the Department of English Philology at the Catholic University of Lublin, Poland, from September 1995 to May 1996, and I am very grateful for all the interesting and inspiring discussions conducted there with late Prof. Edmund Gussmann, and also with Prof. Eugeniusz Cyran. In Gothenburg, the most important support has come from my supervisors: the late Prof. Roman Laskowski, Prof. Nadezjda Zorikhina Nilsson and Dr. Olof Paulsson, who have always been very supportive and encouraging during the various stages my work. I also want to thank the department's leadership warmly for helping to create good working conditions towards the end of this process. Many other colleagues have read parts of the manuscript and engaged in discussions about different problems, and without going into names and details, I want to express my sincere gratitude to all of you who in different ways have helped me to carry this project through to the end, including supervisors, colleagues, friends and family.

Preface

The aim of the present study is to look into one of the scarcely-investigated problems of Slavic languages, namely the vowel–zero alternations in prepositions: to detect, describe and compare the patterns in these alternations in the three major West Slavic languages, Czech, Polish and Slovak. The study springs from a combination of subjects in which I have an interest, namely Czech, Polish, and Slovak linguistics, comparative Slavic linguistics, phonology, morphology and corpus linguistics.

All Slavic languages exhibit vowel–zero alternations to a greater or lesser extent. Such alternations may be internal, within roots or stems, or may occur in prefixes and prepositions, but also, for instance, in adverbs and inflectional endings. The most common case, and also the most frequently discussed in the literature, is that of a vowel which is present in the last syllable of the stem when no suffix is present, while the same vowel is absent as soon as a suffix containing a vowel is added.

Final vowel–zero alternations in prepositions occur in all Slavic languages except Slovenian. However, the alternations in prepositions are rarely touched upon in the literature, much less systematically investigated. The closely related subject, alternations in prefixes, has however been paid far more attention.

In some of the Slavic languages, e.g. in Polish, the circumstances under which a vowel surfaces in consonant-final prepositions are fairly clear. In others, e.g. in Czech, they are rather unclear. As this is a significant obstacle for foreign students, I needed to find out more precisely what regularities could be traced in order to be able to present an appropriate grammatical description of this issue to non-native speakers. When I also encountered the statement “si quelqu’un entreprenait une étude distributionnelle synchronique pour les prépositions semblable à celle que j’ai présentée supra pour les préfixes, alors le résultat en termes de distribution complémentaire serait nul”, by which Scheer (1996: 111) claims that there is absolutely no phonological regularity in vowel–zero alternations in Czech prepositions, this was rather opposed to my intuition, and I therefore began to study the limited literature on the subject. In order to be able to investigate the problem objectively, I started to collect Czech text files for a corpus. I also started to orientate myself in corpus linguistics and suitable concordancing and other programs. In a pilot study in the 1990s I gained data that very much suggested that there are significant regularities to be found. Because I have a profound interest in all the West Slavic languages and I suspected interesting differences between these languages, I decided to make the study comparative and to include all the three major West Slavic languages, i.e. Polish, Slovak and Czech.

Contents

Abstract	5
Acknowledgements	7
Preface	9
Contents	11
Abbreviations and Symbols	17
1. Introduction	19
1.1. Background.....	21
1.2. Motive.....	22
1.3. Aim	23
1.4. Method	23
1.5. The Corpora.....	24
1.6. Programming.....	26
1.7. The West Slavic Languages	27
1.8. Structure of the Study.....	28
2. Phonological Background	31
2.1. Recent Monographs.....	31
2.2. Syllabification	32
2.3. Stress.....	32
2.4. Length.....	33
2.5. Phonemes	34
2.5.1. <i>Polish Phonemes</i>	34
2.5.2. <i>Slovak Phonemes</i>	36
2.5.3. <i>Czech Phonemes</i>	37
2.6. Palatalisation.....	39
2.7. Voicing and Devoicing.....	39
3. Vowel–Zero Alternations	41
3.1. Historical Background.....	41
3.2. Stem-Internal Alternations	43
3.3. Prefix-Final Alternations.....	44
3.4. Preposition-Final Alternations	44
4. Corpora and Method	48
4.1. The Polish Corpus.....	48
4.2. The Slovak Corpus.....	49
4.3. The Czech Corpus.....	49

4.4. Extraction and Analysis of Data.....	50
4.5. Quality of Input Data.....	51
4.6. Corpus Terminology.....	52
4.7 Notational Conventions.....	53
5. General Survey of Data.....	55
5.1. Types of Prepositions.....	57
5.1.1. <i>Asyllabic Prepositions</i>	57
5.1.2. <i>Syllabic Prepositions</i>	57
6. The Polish Data.....	59
6.1. The Corpus Data.....	59
6.1.1. <i>Syllabic Prepositions</i>	60
6.1.1.1. bez(e).....	60
6.1.1.2. nad(e).....	61
6.1.1.3. od(e).....	62
6.1.1.4. pod(e).....	63
6.1.1.5. przed(e).....	64
6.1.1.6. przez(e).....	65
6.1.1.7. spod(e).....	65
6.1.1.8. sprzed(e), znad(e).....	66
6.1.2. <i>Asyllabic Prepositions</i>	66
6.1.2.1. w(e).....	66
6.1.2.2. z(e).....	73
6.2. The Literature.....	83
6.2.1. <i>Dictionaries</i>	83
6.2.1.1. beze.....	86
6.2.1.2. nade.....	86
6.2.1.3. ode.....	87
6.2.1.4. pode.....	88
6.2.1.5. przede.....	89
6.2.1.6. przeze.....	90
6.2.1.7. spode.....	90
6.2.1.8. sprzede.....	91
6.2.1.9. znade.....	92
6.2.1.10. we.....	92
6.2.1.11. ze.....	97
6.2.2. <i>Language Correctness</i>	101
6.2.3. <i>Grammar</i>	107
6.2.4. <i>Linguistics</i>	112
6.3. Summary of Polish Data.....	119
6.3.1. <i>Phonological Vocalisations</i>	120
6.3.2. <i>Lexical Vocalisations</i>	120

7. The Slovak Data	125
7.1. The Corpus Data	125
7.1.1. <i>Syllabic Prepositions</i>	126
7.1.1.1. bez(o).....	126
7.1.1.2. cez(o).....	128
7.1.1.3. nad(o).....	129
7.1.1.4. od(o).....	130
7.1.1.5. pod(o).....	131
7.1.1.6. pred(o).....	132
7.1.1.7. spred(o).....	133
7.1.1.8. spod(o), znad(o).....	133
7.1.1.9. navzdor(y).....	133
7.1.1.10. skrz(e).....	134
7.1.2. <i>Asyllabic Prepositions</i>	136
7.1.2.1. k(u).....	136
7.1.2.2. s(o).....	141
7.1.2.3. v(o).....	147
7.1.2.4. z(o).....	153
7.2. The Literature	157
7.2.1. <i>Dictionaries</i>	158
7.2.1.1. bezo.....	158
7.2.1.2. cezo.....	159
7.2.1.3. nado.....	160
7.2.1.4. odo.....	160
7.2.1.5. podo.....	161
7.2.1.6. predo.....	162
7.2.1.7. spodo.....	163
7.2.1.8. spredo.....	163
7.2.1.9. znado.....	164
7.2.1.10. navzdory.....	164
7.2.1.11. skrze.....	164
7.2.1.12. ku.....	165
7.2.1.13. so.....	168
7.2.1.14. vo.....	170
7.2.1.15. zo.....	173
7.2.2. <i>Language Correctness</i>	175
7.2.3. <i>Grammars</i>	177
7.2.4. <i>Linguistics</i>	184
7.3. Summary of Slovak Data	189
7.3.1. <i>Phonological Vocalisations</i>	190
7.3.2. <i>Vocalisations by Analogy</i>	191
7.3.3. <i>Lexical Vocalisations</i>	192

8. The Czech Data	199
8.1. The Corpus Data.....	199
8.1.1. <i>Syllabic Prepositions</i>	200
8.1.1.1. bez(e)	200
8.1.1.2. nad(e).....	203
8.1.1.3. od(e).....	204
8.1.1.4. pod(e)	206
8.1.1.5. před(e).....	207
8.1.1.6. přes(e).....	209
8.1.1.7. zpod(e)	210
8.1.1.8. zpřed(e)	210
8.1.1.9. poblíž(e)	210
8.1.1.10. skrz(e).....	211
8.1.1.11. díky(y)	211
8.1.1.12. krom(ě).....	211
8.1.1.13. navzdor(y).....	212
8.1.2. <i>Asyllabic Prepositions</i>	212
8.1.2.1. k(e)/ku	213
8.1.2.2. v(e)	231
8.1.2.3. z(e)	248
8.1.2.4. s(e).....	259
8.2. The Literature.....	259
8.2.1. <i>Dictionaries</i>	259
8.2.1.1. beze.....	260
8.2.1.2. nade	261
8.2.1.3. ode	262
8.2.1.4. pode	263
8.2.1.5. přede.....	264
8.2.1.6. přese.....	265
8.2.1.7. zpode.....	266
8.2.1.8. zpřede.....	266
8.2.1.9. poblíže	266
8.2.1.10. skrze.....	266
8.2.1.11. díky	267
8.2.1.12. kromě.....	267
8.2.1.13. navzdory	267
8.2.1.14. ke/ku	267
8.2.1.15. se	269
8.2.1.16. ve.....	269
8.2.1.15. ze	270
8.2.2. <i>Language Correctness</i>	271
8.2.3. <i>Grammars</i>	274

8.2.4. <i>Linguistics</i>	282
8.3. Summary of Czech Data	288
8.3.1. <i>Phonological Vocalisations</i>	289
8.3.2. <i>Vocalisations by Analogy</i>	296
8.3.3. <i>Lexical Vocalisations</i>	297
9. Phonological Analysis	302
9.1. The Representation of V-Ø Alternations.....	302
9.2. First Vocalisation Principle.....	307
9.2.1. <i>Assimilation of Tone</i>	317
9.2.2. <i>Assimilation of Place of Articulation</i>	318
9.2.4. <i>OCP and Syllabic Prepositions</i>	319
9.3. Second Vocalisation Principle	319
9.4. Third Vocalisation Principle	325
9.5. Questions for Further Investigation	330
10. Conclusions	331
Corpus Texts	333
Polish Corpus Texts.....	333
Slovak Corpus Texts	334
Czech Corpus Texts	338
References	341
Sammanfattning på svenska	351

Abbreviations and Symbols

A.	accusative case
C	consonant
C.Sl.	Common Slavic
CNK	Czech National Corpus (Český národní korpus)
Cz.	Czech
D.	dative case
f.	feminine gender
G.	genitive case
GP	Government Phonology
I.	instrumental case
L.	locative case
m.	masculine gender
N	nucleus
N.	nominative case
n.	neuter gender
O	onset
Pl.	Polish
pl.	plural
R	rhyme
S	sibilant
sg.	singular
Sk.	Slovak
SNK	Slovak National Corpus (Slovenský národný korpus)
V	vowel
V.	vocative form
[]	phonetic representation
/ /	phonological representation
{ }	morphemic representation
< >	orthographic representation
→	direction of licensing
1x	once
2x	twice
3x	three times
	etc.

In tables:

- vocalisation occurs in the collocation;
- vocalisation occurs only in the collocation(s);
- (•) vocalisation occurs optionally in the collocation;
- ((•)) vocalisation occurs optionally and rarely in the collocation;

- † vocalisation in the collocation is obsolete;
- sem. vocalisation occurs only in an expression that differs semantically from the same collocation with the non-vocalised form of the preposition;
- phr. vocalisation occurs only in a specific longer fixed phrase;
- R the collocation is used in regional speech varieties.

1. Introduction

Alternations between two shapes of the same morpheme, only differing with respect to the presence or absence of a vowel that appears in one of the forms but not in the other, can be found in all Slavic languages. Such alternations are commonly referred to as vowel-zero alternations. The conditions for the distribution of such alternating forms of prepositions in the three major West Slavic languages constitute the matter for investigation in the present work.

Vowel-zero alternations may be of several different types. They may, however, be divided into two main categories: alternations which occur morpheme-internally, or more precisely non-finally, and alternations which occur morpheme-finally.

The non-final alternations occur in roots, as well as in suffixes and prefixes. An example of a root-internal vowel-zero alternation is found in the Czech noun **pes** *dog*, which drops the vowel **-e-** as soon as a desinence is added. We therefore have the N.pl. **psi** with the desinence **-i**. Hence, the overtly alternating allomorphs are /pes/ ~ /ps/.

An example of a suffix with a non-final vowel-zero alternation is found in the Slovak diminutive suffix **-ec**, e.g. in the noun **chlapec** *boy*, where the **-e-** is dropped before any desinence, giving the N.pl. **chlapci**. Hence, the overtly alternating allomorphs of the diminutive suffix are /ec/ ~ /c/.

Internal alternations in prefixes are much more rare, but an example is found in the Polish verbal prefix **wez-**, e.g. in the perfective verb **wezbrać** *to well up*, where the vowel is dropped in the corresponding imperfective verb **wzbierać** with the same meaning. Hence, the overtly alternating allomorphs are /vez/ ~ /vz/.

Instances of morpheme-final vowel-zero alternations are mainly found in prepositions and prefixes, and only sporadically in other types of morphemes. Prefix-final vowel-zero alternations occur in derived lexemes, e.g. the alternation between the allomorphs /v/ and /ve/ in the Czech derived verbs **vjede** *drives in* (from **jede** *drives*) versus **vejde** *walks in* (from **jde** *walks*). Such derivations tend to be rather difficult to analyse synchronically, as they have often been lexicalised at some more or less remote stage in the history of the specific language. Therefore, they may be a result of phonological principles which no longer apply, but which did apply to the language at the specific earlier stage in its development when the lexeme was established; however, they may also be influenced by later analogies and simplifying unifications of stems.

The types of vowel-zero alternations exemplified so far, i.e. non-final alternations in roots and suffixes, as well as final alternations in prefixes, have been thoroughly investigated, described and discussed for the vast majority of Slavic languages, and they will therefore not be discussed any further in the

present work. They also have in common that they are traditionally considered to be phonologically driven; cf. e.g. Gussmann (2007) for Polish, Rubach (1993) for Slovak and Scheer (1996) for Czech. Whether or not that is so, is of course an extremely interesting question, but it would lead us much too far to undertake that discussion at present.

Quite on the contrary, though, preposition-final vowel-zero alternations, as, for instance in the Slovak collocations **vo vode** *in the water* vs. **v škole** *at school* (which could also be compared to the Czech **ve vodě**, **ve škole** and the Polish **w wodzie**, **w szkole**), have not been frequently discussed at any length in the literature on Slavic phonology. The traditional standpoint is that these alternations, as well, are phonologically driven, but, as we shall see in the following chapters, the traditional explanations no longer hold for the phenomena that can be observed with regard to prepositions in the modern Slavic languages. As a reaction to this, Scheer (1996: 111) claims that preposition-final vowel-zero alternations are not synchronically driven by phonology and that an investigation of the issue would show that no complementary distribution of the forms can be found.

The range of the present investigation is restricted to the three major West Slavic languages, Polish, Slovak and Czech, but this type of preposition-final vowel-zero alternation occurs in all Slavic languages except for Slovene. All the rest of the Slavic languages have prepositions that exhibit alternations between a shorter and a longer form, also known as the non-vocalised and the vocalised form, respectively. The conditions, however, under which each of the two forms is used, differ from one language to the other, as can easily be observed in Slovak, Czech and Polish examples above.

Because vowel-zero alternations in prepositions occur in very occasional collocations, it is most plausible that they are triggered mainly by synchronic principles. These synchronic principles, however, do not necessarily coincide with the phonological principles that have traditionally been set up for Slavic languages in general. For these reasons, I expect vowel-zero alternations in prepositions to offer a highly interesting field of research. I have chosen to investigate these alternations rather thoroughly in the three major West Slavic languages by taking a substantial portion of corpus data as the point of departure for the investigation.

Finally, it should be mentioned that some morpheme-final alternations in other types of morphemes are to be found, for instance in the Czech present tense desinence of the 1st person plural, as in e.g. **neseme** *we carry* with the alternate form **nesem**, or in Czech adverbs such as e.g. **zase** *again; on the other hand* and **více** *more* with the alternate form **zas** and **víc**. Such final alternations are also found in some Serbo-Croatian adjectival desinences, e.g. the A.G.sg. **-og/-oga**, the L.D.sg. **-om/-ome/-omu** or the L.D.I.pl. **-im/-ima**. The distribution of such alternating forms is, however, rather free, and the major factors determining the choice of allomorph are stylistic, rhythmical or

euphonic (Mønnesland 2002: 52, Diklić 1979: 108). Such alternations are therefore probably not of any major interest from a phonological point of view.

1.1. Background

My interest in phonology was awakened many years ago by Professor Roman Laskowski. The issues attracting my attention at that point were the Ukrainian [u]-[w]-[v]¹ and [i]-[j] alternations. The traditional analyses did not satisfy my curiosity, so I began to study different contemporary trends in phonology. However, it was not until I ran across Government Phonology (GP) as presented by Charette (1991), that I saw all the pieces falling into their places. This framework fascinated me from the very beginning. In her study, Charette deals with vowel-zero alternations, mainly in French material, and that inspired me to embark on a study of Slavic vowel-zero alternations within the framework of GP.

During a stay at the Catholic University of Lublin, Poland, I had the opportunity to become thoroughly acquainted with GP, which at that time was the prevailing phonological school of thought in the University's English department. I then also made a first small investigation of vowel-zero alternations in Czech prepositions on a corpus of about one million words (Nilsson 1996). At that point, it was intended as only a part of a larger general work about different types of vowel-zero alternations and syllable structure in Czech. However, I became so inspired by the problem of vowel-zero alternations in prepositions, that after some time, it seemed natural to concentrate the thesis on this problem alone.

At the same time, I was intrigued by the fact that Slovak and Polish exhibit patterns of vowel-zero alternations in prepositions that are significantly different from the Czech patterns (Nilsson 1998). From the traditional point of view, this was highly unexpected, as more or less the same phonological principles were supposed to be responsible for the vowel-zero alternations in all Slavic languages. I therefore decided to include all three major West Slavic languages in the investigation and to make the study corpus-based and comparative.

However, the corpora of the West Slavic languages that were available in the 1990s were very small. At the same time, ordinary computers could not handle very large amounts of data quickly. It would take several minutes to search for an expression in a corpus of one million words. Hence, two contradictory obstacles stood in the way: too little text to give enough input data for my investigation, but too much text to be processed in a reasonable amount of time.

¹ An attempt at an interpretation of this type of alternations is made in Cyran & Nilsson (1998).

The situation has improved a lot during the last ten years. Faster computers make it possible to search a corpus of 50 million words in less than a minute. In the Czech and Slovak Republics, large national corpora have been established and are freely accessible on the Internet. Also, many Czech and Slovak newspapers and magazines have open archives on the Internet and many e-books are freely available in Czech and Slovak e-libraries. Only in Poland is the amount of freely accessible e-resources, such as e-books, newspaper archives and text corpora, somewhat more restricted. However, today it is not a problem to acquire enough input data for this type of investigation, nor to process these data.

1.2. Motive

The major motive of this investigation is the quite appalling lack of thorough descriptions and precise information on the vowel-zero alternations in prepositions in the West Slavic languages. The Polish situation has been described systematically, but very briefly, by Klemensiewiczówna (1950), Klemensiewicz-Bajerowa (1952, 1955), Bańko (2002: 114f.) and Gussmann (2007: 234ff.). However, quite a few interesting and important details lack thorough investigation and discussion. On the other hand, a number of other Polish authors give a strikingly inadequate, not to say erroneous picture of the issue. The situation with regard to Slovak and Czech is far more unclear. The situation in Slovak has been partially discussed by Peciar (1947), Sabol (1966), Oravec (1968), Rubach (1993), and Vojteková (2011), while the situation in Czech has been briefly discussed by Ryba (1934), Haller (1934) and Kučera (1982, 1984), but no attempt at a systematic investigation and description has been made, and comments on the issue that can be found scattered in other works are often accompanied by very unsatisfactory, or sometimes even false conclusions.

It should also be added that due to the lack of corpora at the time when most of these texts were written, all of them are primarily based on occasional excerpts and, above all, on the native authors' intuition and sense of their mother tongue.

An investigation of the actual present day situation with regard to vowel-zero alternations in prepositions is very much needed by foreign learners of West Slavic languages, but sometimes also by native speakers, who have doubts about the appropriateness of their own usage. This investigation could also serve as a source of information about the actual preferences in the modern usage of vocalised forms of prepositions in present day Polish, Slovak and Czech, and the data it contains might thereby serve as arguments in language correctness discussions, as well as providing a reference point for authors of future normative grammars of these languages.

1.3. Aim

The primary aim of this thesis is to track down and formulate, as precisely as possible, the relevant principles that govern the choice of the vocalised or the non-vocalised form of the prepositions in each of the three languages investigated, as well as to present lists of exceptions to these principles and to describe the status of such exceptions.

A secondary aim of the study is to compare the phonological principles governing the vocalisation of prepositions in the three languages to each other. Hopefully, some general observations can be made about the principles governing these vowel-zero alternations, and about the types of structures that trigger these alternations.

An additional aim of this investigation will be to test the suitability of the chosen framework for analysing this kind of data, to see whether it can serve as a powerful phonological tool when describing a completely new set of data to which it has not been applied previously. The study will also provide the framework with comparative data from the field of vowel-zero alternations, which have already been investigated within this framework with regard to a number of other languages, e.g. French (Charette 1991), Arabic (Kaye 1990) and Japanese (Yoshida 1990, 1991).

1.4. Method

The investigation is based on data extracted from computational text corpora of each of the three languages. All instances of each form of the alternating prepositions are retrieved computationally. The output data are basically of two types: firstly, purely statistical information, i.e. the number of occurrences of each form of the prepositions in different phonological contexts, and secondly, alphabetic listings of every single occurrence in its context. For reference and comparison, additional data is sometimes also retrieved from other freely accessible national corpora.

For each pair of prepositions, one vocalised and one non-vocalised, two sets of data are retrieved, one set of data for each form of the preposition. The two sets of data are compared and subdivided with regard to the initial part of the following word. The relevant phonological context is considered to be the initial consonant(s) before the first realised nucleus of the post-prepositional word, which in most instances is equivalent to the first vowel of the word following the preposition.

The two sets of data are first compared purely statistically. In some contexts, clear tendencies can be traced based merely on this statistical output, as the distribution is 100% versus 0% or extremely close to these figures. An attempt can then be made to formulate a phonological principle responsible for this behaviour without any deeper analysis of the corpus data. In many

other cases the statistical output is not so straight-forward, and all the contextual examples in the corpus data have to be closely scrutinised in order to reveal the patterns and try to pinpoint the principles at work. In order to do so, in each set of data where both forms of a preposition occur in the same phonological context, all the instances of the form of the preposition which is least frequent are investigated manually.

When all the data have been thoroughly analysed, a final attempt is made to formulate the general phonological principles for the vowel–zero alternations in all the prepositions in each of the languages by means of the GP framework. An attempt is also made to systematically describe the similarities and differences between the three phonological systems and to show how these differences are responsible for the differences in the actual realisations of vowel–zero alternations in the three languages investigated.

Throughout the work, all cases of non-phonologically conditioned vocalisations are systematically singled out in order to make the phonological picture of the problem investigated as clear as possible. All such exceptions to the phonological principles that are formulated are also systematically listed.

1.5. The Corpora

For the three languages investigated, there now exist well-established, comprehensive, web-based national corpora: *Český národní korpus* (CNC)², *Slovenský národný korpus* (SNK)³ and *Narodowy korpus języka polskiego* (NKJP)⁴. However, for various reasons these corpora could not be used for this investigation. Firstly, the type of searches involving a string of one word plus the beginning of the following word could not be carried out in a desirable way with the available tools. Secondly, the large amounts of data that are required for this investigation could not be conveniently retrieved from the corpora. Thirdly, the retrieved data could not be saved locally in a convenient format allowing further division and sorting of the material. Finally, these corpora have developed over time, and were not as easily available when this investigation was started as they are today. In fact, the Polish National Corpus has only been available since the end of 2012, and its three predecessors *Korpus IPI PAN*⁵, *PELCRA*⁶, and *Korpus Języka Polskiego Wydawnictwa Naukowego PWN*⁷, offered even more restricted possibilities to search and save retrieved

² Homepage at <<http://ucnk.ff.cuni.cz/>>, last accessed 2014-08-24.

³ Homepage at <<http://korpus.juls.savba.sk/>>, last accessed 2014-08-24.

⁴ Homepage at <<http://www.nkjp.pl/>>, last accessed 2014-08-24.

⁵ Homepage at <<http://korpus.pl/>>, last accessed 2012-09-18.

⁶ Homepage at <<http://pelcra.pl/>>, last accessed 2012-09-18.

data than their Czech and Slovak counterparts.

Because of the restricted possibilities of working with the existing national corpora, it became necessary to build corpora on a local computer and work on them with sophisticated concordance tools in order to facilitate searches for the required strings of words and parts of words, to retrieve very large quantities of data in a relatively short time, and also to save the data extracted for further processing. Therefore input data was collected from freely accessible texts on the Internet.

The sizes of the corpora of the different languages are shown in the following table.

	Polish	Slovak	Czech
EURG	172,038	24,151	28,089
FACT	291,687	1,509,845	0
FICT	493,538	8,690,847	5,236,510
FICTR	0	0	760,574
PRLM	4,012,412	0	0
PRSS	908,300	20,652,094	51,810,132
Total	5,877,975	30,876,937	57,835,305

Table 1. Size of the corpora: number of tokens.⁸

EURG: European Union regulations. These are source texts for my own translations into Swedish. The texts are located at

<<http://ec.europa.eu/enterprise/tris/>>.

FACT: Factual texts available at open access e-libraries in Poland and Slovakia. These e-libraries are located at

<<http://www.twojebook.net/>>, <<http://zlatyfond.sme.sk/>>,

<<http://knihy.saleziani.sk/>>.

FICT: Fiction available at open access e-libraries in Poland, Slovakia and the Czech Republic. These e-libraries are located at

<<http://exlibris.biblioteka.prv.pl/>>, <<http://zlatyfond.sme.sk/>>,

<<http://e-knihovna.cz/>>.

FICTR: Fiction in translation from a foreign language located at the same e-library as above.

PRLM: Transcripts of the proceedings of the Polish Parliament from January to September 2006. These texts are located at

<<http://www.sejm.gov.pl/>>.

PRSS: Press. These texts were gathered from open access web archives

⁷ Homepage at <<http://korpus.pwn.pl/>>, last accessed 2012-09-18.

⁸ Tokens consisting of digits are not included.

of several Czech, Slovak and Polish newspapers and magazines in the late 1990s. Some of these websites no longer exist, but a list of titles and dates of issue, as well as the addresses of the still existing sites, are found in the bibliography.

The three individual language corpora are of noticeably different sizes. The largest effort has been made to compile Czech material, as a pilot study clearly showed that this language exhibits the most intriguing vowel–zero alternations in prepositions, whereas the Polish language seemed to exhibit the least complicated system.

Because of the need for large amounts of text, the most easily accessible texts in the simplest available text format were chosen. For this reason, the Czech and Slovak corpora consist to a major extent of texts from the Czech and Slovak press that were located in the open archives of various newspapers and magazines, whereas the Polish corpus consists mainly of transcripts of parliamentary proceedings. However, for the purposes of this study, the type of text is likely to be of negligible significance, as the object of investigation is the standard written variety, and the principles responsible for the alternations investigated are unlikely to differ significantly from one kind of text to another. It might even be that newspaper texts and parliamentary transcripts better reflect the neutral present day standard literary usage than does fiction, as in fiction there is a larger risk that the texts may bear traces of the authors' individual usage, as the acceptability of a personal style is considerably higher in fiction.

The quality of the texts included in the corpus is quite good, but of course they contain a certain number of traditional typing errors, e.g. **ze** for **za**, as well as some computational errors, e.g. division of words such as the Czech **divi ze** instead of **divize** *division*. Such errors, of course, make the total number of occurrences of the preposition **ze** slightly higher, but all contradictory data are investigated, hence such errors will finally be detected, if they contradict other stronger tendencies in the corpus material.

1.6. Programming

In order to retrieve the data needed for this investigation from the text corpora, I have chosen the corpus analysis programme WordSmith Tools (WS)⁹, which allows the user to define the search parameters in a very sophisticated way, to extract large amounts of occurrences together with a freely selected amount of context, and subsequently to sort these occurrences in their context according to an elaborate set of parameters.

The programme also allows you to save your search results for later use, as

⁹ Detailed information at <<http://www.lexically.net/wordsmith/>>, accessed 2014-08-24.

well as your selected sortings and edited sets of occurrences.

All these functions of WS are much more flexible than the tools offered by the web-based copora on the Internet. The possibility to save the search results is particularly important, as one only has to carry out each search once, and later it is possible to work with exactly the same set of output data whenever it is needed.

The fact that the data is processed on a local computer also significantly reduces the time needed to carry out the searches.

1.7. The West Slavic Languages

Traditionally, Slavic linguists consider the group of West Slavic languages to consist of the five languages¹⁰ known as Polish, Czech, Slovak, Upper Sorbian and Lower Sorbian¹¹. However, in this study only the three major West Slavic languages¹² will be investigated.

The phonetic and phonological characteristics of the three major West Slavic languages will be given in a separate chapter. Here, only some brief sociolinguistic characteristics will be provided.

The Polish and Slovak standard languages are rather homogeneous, with only quite small differences between the written standard language and the everyday spoken standard language. Furthermore, in Poland dialects are not in frequent use, with the exceptions of the Silesian, Kashubian and Góral¹³ areas. Most of these speakers are however able to switch quite freely between the regional variety and standard spoken Polish.

In Slovakia, dialects have a stronger position. People who speak a dialect

¹⁰ Some scholars would include more than five languages. They would, above all, add Rusinian, spoken in Vojvodina, but also Kashubian, spoken in northern Poland. The latter would, however, be considered a dialect by others. The status of a few other Czech, Slovak and Polish dialects have also at times been vividly discussed.

¹¹ The two Sorbian languages are also referred to by many scholars as *Lusatian* (in Polish: *łużycki*, in Czech: *lužický*), and especially in German also as *Wendisch*.

¹² Taking into account data from many different sources, it seems quite evident that Polish has substantially more than 40 million speakers, Czech more than 11 million and Slovak more than 5 million. Upper Sorbian has only some thousands of speakers. Probably all of them are bilingual and most of them have German as their stronger language, i.e. they have a rather passive knowledge of Upper Sorbian. For Lower Serbian the figures are even smaller, with only a very few active speakers.

¹³ The Silesian regional variety is spoken in an area in and around the metropolitan conglomerate of Katowice and its sister cities. The Kashubian regional variety is spoken in a rural area to the west of the city of Gdańsk. The Góral regional variety is spoken in the mountain area of the High Tatras, in and around the city of Zakopane.

are in general not readily able to switch as easily to the standard spoken language, but there is a strong tendency in the younger generation to abandon the dialects in favour of the spoken standard. In some regions, however, especially in the eastern part of the country, a regional standard is often used instead of the genuine dialects. These regional standards are much closer to the literary standard, but retain some of the general characteristics of the dialects in the region.¹⁴

Czech, on the contrary, exhibits a different situation. It is commonly used in several different varieties. First, there exists a dichotomy between the literary standard (codified norm) and the common colloquial language (usage), which however has its own unwritten standard (Townsend 1990, Sgall & Hronek 1992, Gammelgaard 1997). These two varieties are traditionally referred to as Literary Czech (in Czech *spisovná čeština*) and Common Czech (in Czech *obecná čeština*). In addition to this dichotomy, the traditional dialects are quite alive in significant parts of the country, even though the younger generations are more and more adjusting to the less marked variety of the common colloquial language. Also, the colloquial standard is not homogeneous all over the country. The variety referred to as Common Czech is mainly used in the western part of the Czech Republic, i.e. in the Bohemian part, whereas in the Moravian and Silesian parts¹⁵, slightly different regional standards have evolved, preserving some of the characteristics of the dialects in the area.¹⁶ The subject of this study is however the written standard language as it is found in the press, administrative texts and fiction, but it cannot be excluded that certain individual occurrences in the corpus involve vocalisations of prepositions which are more or less based on a non-standard usage.

1.8. Structure of the Study

In Chapter 2, a concise introduction is given to the phonological system of each of the three major West Slavic languages. The extent of these surveys is restricted to facts which are indispensable in order for the reader to follow the subsequent investigation of the vowel–zero alternations in West Slavic prepositions.

¹⁴ Typical eastern Slovak features include, above all, penultimate stress and lack of long vowels. Both these features also happen to occur in Polish.

¹⁵ The eastern part of the Czech Republic is constituted by Moravia (south and central parts) and Silesia (northern parts) with their respective main cities Brno and Ostrava.

¹⁶ Typical Moravian features include, above all, a different realisation of some of the long vowels and diphthongs as well as assimilation processes concerning the obstruent consonants. Typical Silesian features are, in addition to the Moravian features, penultimate stress and lack of long vowels, as in Polish.

In Chapter 3, the phenomenon of vowel–zero alternations in Slavic languages will be introduced in more detail and discussed from a general and historical point of view in order to offer a background to the present investigation. Stem-internal, prefix-final and preposition-final alternations will be discussed in order to give a more complete picture of the phenomenon.

In Chapter 4, the corpora and the method used in order to extract the data from the corpora will be presented in more detail.

In Chapter 5, some introductory and more general observations drawn from the corpus data will be presented. Some overall statistical data from the individual languages will be presented and compared to each other. Further, a general division of the prepositions into primary and secondary as well as into syllabic and asyllabic will be presented, as different principles seem to be at work with respect to these different types of prepositions.

In Chapters 6, 7 and 8, the conditions for the vowel–zero alternations in prepositions in each of the three major West Slavic languages will be investigated in detail. First, the relevant corpus data on prepositions exhibiting vowel–zero alternations will be presented. The data will be systematically investigated and analysed, in search of phonological regularities and in order to distinguish the phonological factors triggering the alternations from other types of triggering factors. Then the available literature on the subject will be summarised and discussed. Details will be drawn from various dictionaries, reference grammars, articles and monographs. The sets of data from the various sources are subsequently compared and discussed in order to track down the exact factors triggering vocalisation of prepositions in the three languages.

In Chapter 9, the phonological contexts triggering vocalisation according to the findings in the preceding chapters are analysed phonologically and the systems of the three individual languages are compared to each other in order to find common patterns and, as far as possible, to unify the sets of phonological principles that can explain the relevant vowel–zero alternations in each of the languages in one coherent model. The phonological analysis is carried out within the theoretical framework of Government Phonology (GP), which will at the same time be introduced in sufficient detail to enable the reader to follow the analysis and discussion.

2. Phonological Background

In order to facilitate understanding of the following discussion for those not familiar with the languages investigated, this chapter will briefly present some fundamental details about the phonological systems of the three languages, above all the basic correlations between orthography, phonemes and phonetic realisations. For obvious reasons, this presentation is forced to be rather schematic. For more detailed and nuanced presentations the reader is referred to one of the existing monographs dealing with the phonology of the individual languages.

Phonological notations are given between slashes, e.g. /dúm/, whereas phonetic transcriptions are given between square brackets, e.g. [du:m]. Where needed for clarity, notations in standard orthography are given between angle brackets, e.g. ⟨dúm⟩. Elsewhere they have not been marked in any special way.

Throughout the text, phonetic transcriptions are given according to the Handbook of the IPA (1999), whereas phonological transcriptions are given according to the Slavic tradition, which means that the diacritic signs used in the standard orthography are also used in the phonological transcription. For Polish, the orthographic digraphs are rendered phonologically by the corresponding dominating graphemes found in other Slavic alphabets, i.e. /š/ for Polish ⟨sz⟩, etc. The digraph ⟨ch⟩, found in all three major West Slavic languages, is rendered both phonologically and phonetically by the IPA sign, i.e. /x/, as well as [x].

2.1. Recent Monographs

More recent monographs on Polish phonology have been published by Wierzchowska (1980), Bethin (1992), Szpyra (1995), Dukiewicz (1995), Sawicka (1995), Strutyński (1995), Ostaszewska & Tambor (2000), Wiśniewski (2001), Kraska-Szlenk (2003), and Gussman (2007), among others. The issue of vowel-zero alternations in prepositions is touched upon very briefly by some of these authors, but only Gussman (2007: 234ff) discusses the issue at any length. There are also two articles (Klemensiewiczówna 1950, Klemensiewicz-Bajerowa 1955) that deal with this issue.

Monographs on Slovak phonology have been published in more recent years mainly by Král & Sabol (1989), Sabol (1989), Rubach (1993) and Bujalka, Baláž & Rýzková (1998). In some of these works, vowel-zero alternations in prepositions are mentioned, but not discussed at any length. Only an article by Sabol (1966) discusses the issue briefly.

Modern monographs on Czech phonology are surprisingly few: Kučera (1961), Kučera & Monroe (1968), Vachek (1968) and Scheer (1996). Other works, mainly dealing with phonetics, but partly concerned with phonology, have been published by Hála (1962, 1967), Pettersson (1979), Krčmová (1984),

Romportl (1985), Mluvnice'86, Mluvnice'95, Palková (1994) and Hůrková (1995). However, none of these works investigates vowel–zero alternations in prepositions. In comparison with Polish, articles on Czech phonology are also surprisingly few. However, four such articles (Haller 1934, Ryba 1934, Kučera 1982, 1984) actually deal with vowel–zero alternations in prepositions.

2.2. Syllabification

In Polish, each phonetic syllable is arranged around a vowel constituting the nucleus of the syllable. In Slovak and Czech, however, the sonorant consonants /l/ and /r/ may also constitute the nucleus of a syllable.¹⁷ When they do, they may also carry word stress.

- (1) Pl. **targ** ['tark] *market*
Sk./Cz. **trh** ['tr̩x] *market*

2.3. Stress

Polish has predictable stress on the penultimate syllable, e.g. **zarabiamy**¹⁸ *we earn*, although some exceptions do exist.¹⁹

Furthermore, a preposition before a monosyllabic pronoun normally attracts the stress from the pronoun, which suggests that the preposition and the pronoun are treated as a single phonological unit.

- (2) Pl. **do niej** *to her*, **przed nim** *in front of him*, **dla nas** *for us*, **po co** *for what purpose*

On the other hand, a preposition before a monosyllabic noun does not normally attract the stress from the noun.

- (3) Pl. **do wsi** *to the village*, **przed dom** *in front of the house*, **dla lwa** *for the lion*

¹⁷ Furthermore, in Standard Czech, /m/ functions as the nucleus of an unstressed syllable in the root of the two words **sedm** ['sedm] *seven* and **osm** ['osm] *eight*. Also, in proper names of foreign origin, final /n/ often functions as the nucleus of a syllable, e.g. **Seton** ['si:tŋ].

¹⁸ Where appropriate, stress will be marked by underlining the vowel or sonorant in the stressed syllable.

¹⁹ It is mainly words of foreign origin (e.g. **lingwistyka** *linguistics*) and some verb forms with agglutinated endings (e.g. **robilibyśmy** *we would do*) that have their stress on the third or fourth syllable from the end of the word, according to the literary norm. However, in the colloquial Polish such words are often stressed on the penultimate syllable as well.

However, there exist a substantial number of exceptions to this last principle, and in those cases the preposition and the noun seem to be treated as a phonological unit.

(4) Pl. **na wsi** *in the countryside*, **na dół** *down*, **na noc** *for the night, over night*

In standard Slovak, stress falls on the initial syllable.²⁰ Prepositions are proclitics and they do in some instances attract stress. This is especially so when a preposition precedes a mono- or bisyllabic noun or pronoun. In other contexts, the proclitic prepositions are less likely to attract stress from the following word. These general tendencies can however be overridden by contrastive emphasis (Kačala 1976: 39; Oravec, Bajzíkóvá & Furdík 1988: 182; Král 1988: 166ff.; Kočíš 1990: 85; Navrátil 2001: 54).

(5) Sk. **do dna** *to the bottom*, **do tej ulice** *into that street*,
but **do Bratislavy** *to Bratislava*

In older works, however, it is often claimed that stress must always fall on monosyllabic and vocalised asyllabic prepositions. (Bartek 1944: 135; Stanislav 1953: 289f.)

In standard Czech, stress falls on the initial syllable. Monosyllabic and asyllabic prepositions are proclitics. Monosyllabic and vocalised asyllabic prepositions then in principle attract stress from the following word. If the following word is long or difficult or semantically in focus, stress may sometimes remain on this word for emphasis and contrast (Mašín 1928: 23, Havránek & Jedlička 1981:36, Hürková 1995: 38f.).

(6) Cz. **na stůl** *onto the table*, **do té ulice** *into that street*,
do Bratislavy *to Bratislava*

2.4. Length

Polish does not have phonological vowel length. It does however have phonological consonant length, which occurs in word roots as well as when two identical phonemes occur next to each other. Long consonants only occur between vowels or word-initially.

(7) Pl. **wanna** ['van:a] *bath tub*, **w wodzie** ['v:ɔdzɛ] *in the water*

Standard Slovak and Czech have phonological vowel length.

(8) Cz. **dal** ['dal] *he gave*, **bila** ['bi:la:] *she hit*, **děla** ['jɛla] *cannons*
dál ['da:l] *farther*, **bílá** ['bila] *white*, **dělá** ['jɛla:] *he/she makes*

²⁰ In the eastern regional variety, stress falls on the penultimate syllable, as in Polish.

They also have contrastive consonant length, but occurrences of long consonants are generally less frequent in Slovak and Czech than in Polish, and even less frequent in Czech than in Slovak.

- (9) Cz. **nejasnější** [ˈnejasnejʃi:] *more unclear*
nejjasnějí [ˈnej:asnejʃi:] *most clear*

On the other hand, Slovak has the two long consonant phonemes /r̄/ [r:] ⟨ř⟩ and /l̄/ [l:] ⟨ĺ⟩. Their occurrence is however restricted to the nucleus of a syllable.

- (10) Sk. **mŕtvy** [ˈmɾ:tvɨ] *dead* Cz. **mrtvý** [ˈmɾtvɨ:] *dead*

The rest of the occurrences of phonetically long consonants in Slovak, as well as all the occurrences in Czech, i.e. those not occurring in a nucleus, are the result of two adjacent instances of the same phoneme, e.g. Sk. **oddych** [od:iχ] *recreation, rest*, which consists of the prefix **od-** *from* and the root **dych** *breathing*. The pronunciation of such long consonants is a frequent issue in language correctness discussions in both languages, thus implying an unstable usage.

2.5. Phonemes

2.5.1. Polish Phonemes

The presentation of the Polish phoneme inventory is mainly based on Ostaszewska & Tambor (2000)²¹.

Polish has six vowel phonemes:

Phoneme	/i/	/y/	/e/	/a/	/o/	/u/
Phone	[i]	[i]	[ɛ]	[a]	[ɔ]	[u]
Grapheme	⟨i⟩	⟨y⟩	⟨e⟩	⟨a⟩	⟨o⟩	⟨u⟩, ⟨ó⟩

Traditionally, Polish is also considered to have two nasal vowels /ɛ̃/ and /ɔ̃/, but in recent times, they are more often treated as the nasal diphthongs /eŵ/ and /oŵ/, as the nasal element is pronounced after an oral vowel and adjusts by assimilation to the following phonological context. Hence, Polish can synchronically be considered to have six oral vowel phonemes and one nasal semi-vowel phoneme /ŵ/.

²¹ They also include the phonemes /k̄/ and /ḡ/, which I prefer to interpret as the phoneme combinations /kj/ and /gj/. However, this is of no importance to this investigation.

Phonemes	/e/+/w̃/	/o/+/w̃/
Allophones	[ẽw̃], [ɛw], [ɛm], [ɛn], [ɛŋ], [ɛɲ], [ɛɲ], [ɛ]	[ɔ̃w̃], [ɔw], [ɔm], [ɔn], [ɔŋ], [ɔɲ]
Grapheme	⟨ę⟩	⟨ą⟩

(11) Pl. **zęby** /zew̃by/ ['zɛmbi] *teeth*

Polish has the following consonant phonemes:

Phoneme	/p/	/b/	/f/	/v/	/t/	/d/
Allophones	[p], [pʲ] [b]	[b], [bʲ] [p]	[f], [fʲ] [v]	[v], [vʲ] [f]	[t], [tʲ] [d]	[d], [dʲ] [t]
Grapheme	⟨p⟩	⟨b⟩	⟨f⟩	⟨w⟩	⟨t⟩	⟨d⟩

Phoneme	/s/	/z/	/ś/	/ź/	/ś/	/ż/
Allophones	[s], [sʲ] [z]	[z], [zʲ] [s]	[ɕ] [z]	[z] [ɕ]	[ʃ] [ʒ]	[ʒ] [ʃ]
Grapheme	⟨s⟩	⟨z⟩	⟨ś⟩ ⟨si⟩ ²²	⟨ź⟩ ⟨zi⟩	⟨sz⟩	⟨ż⟩, ⟨rz⟩

Phoneme	/c/	/d͡z/	/ć/	/d͡ʒ/	/č/	/d͡ʒ/
Allophones	[ts] [d͡z]	[d͡z] [ts]	[tɕ] [d͡z]	[d͡z] [tɕ]	[tʃ] [d͡ʒ]	[d͡ʒ] [tʃ]
Grapheme	⟨c⟩	⟨dz⟩	⟨ć⟩ ⟨ci⟩	⟨d͡ʒ⟩ ⟨dzi⟩	⟨cz⟩	⟨d͡ʒ⟩

Phoneme	/k/	/g/	/x/	/m/	/n/	/ń/
Allophones	[k], [kʲ] [g]	[g], [gʲ] [k]	[x], [xʲ] [ɣ]	[m], [mʲ], [mʲ]	[n], [ŋ]	[ɲ]
Grapheme	⟨k⟩	⟨g⟩	⟨ch⟩, ⟨h⟩	⟨m⟩	⟨n⟩	⟨ń⟩ ⟨ni⟩

Phoneme	/r/	/l/	/w/	/j/		
Allophones	[r], [rʲ]	[l], [lʲ]	[w]	[j]		
Grapheme	⟨r⟩	⟨l⟩	⟨ł⟩	⟨j⟩		

²² ⟨si⟩ etc. is written before a following vowel, ⟨ś⟩ etc. before a consonant and also word-finally. When the following vowel is /i/, only one ⟨i⟩ is written, e.g. **sito** /śito/ *sieve*.

2.5.2. Slovak Phonemes

The presentation of the Slovak phoneme inventory is mainly based on Rubach (1993).²³

Slovak has six vowel phonemes:

Phoneme	/i/	/e/	/ä/	/a/	/o/	/u/
Allophones	[i]	[ɛ]	[æ], [ɛ] ²⁴	[a]	[ɔ]	[u]
Grapheme	⟨i⟩, ⟨y⟩	⟨e⟩	⟨ä⟩	⟨a⟩	⟨o⟩	⟨u⟩

Slovak has five long vowels. The long counterpart to /ä/ is the diphthong /ia/.

Phoneme	/i:/	/é/	/ä:/	/ó/	/ú/
Allophones	[i:]	[ɛ:]	[a:]	[ɔ:]	[u:]
Grapheme	⟨i̇⟩, ⟨ý⟩	⟨é⟩	⟨ä̇⟩	⟨ó⟩	⟨ú⟩

Slovak has four diphthongs:²⁵

Phoneme	/ie/	/ia/	/iu/	/uo/
Allophones	[iɛ]	[iä]	[iu]	[uɔ]
Grapheme	⟨ie⟩	⟨ia⟩	⟨iu⟩	⟨ô⟩

Slovak has the following consonant phonemes:

Phoneme	/p/	/b/	/f/	/v/	/t/	/d/
Allophones	[p] [b]	[b] [p]	[f] [v]	[v] [u̇], [f]	[t] [d]	[d] [t]
Grapheme	⟨p⟩	⟨b⟩	⟨f⟩	⟨v⟩	⟨t⟩	⟨d⟩

Phoneme	/s/	/z/	/š/	/ž/	/k/	/g/
Allophones	[s] [z]	[z] [s]	[ʃ] [ʒ]	[ʒ] [ʃ]	[k] [g]	[g] [k]
Grapheme	⟨s⟩	⟨z⟩	⟨š⟩	⟨ž⟩	⟨k⟩	⟨g⟩

²³ However, he does not include a phoneme /j/ alongside the phoneme /i/, as the attachment of the phoneme to a nucleus or not is enough to make the distinction. I agree with that view, but include /j/ here for clarity.

²⁴ [æ], [ɛ] as realisations of /ä/ occur in free variation.

²⁵ Slovak combinations like /ow/, /aw/, /ej/, /aj/ etc. are treated as vowel + consonant. This issue, however, does not influence the present investigation.

Phoneme	/c/	/ɟ/	/č/	/č̣/	/t̥/	/d/
Allophones	[s] [ɟ]	[ɟ] [s]	[tʃ] [dʒ]	[dʒ] [tʃ]	[c] [j]	[j] [c]
Grapheme	⟨c⟩	⟨dz⟩	⟨cz⟩	⟨dž⟩	⟨t̥⟩, ⟨t⟩ ²⁶	⟨ḍ⟩, ⟨d⟩ ²⁶

Phoneme	/x/	/h/	/m/	/n/	/ň/
Allophones	[x] [ɣ]	[h] [x]	[m], [m̥]	[n], [n̥]	[ɲ]
Grapheme	⟨ch⟩	⟨h⟩	⟨m⟩	⟨n⟩	⟨ň⟩, ⟨n⟩ ²⁶

Phoneme	/r/	/ř/	/l/	/l̥/	/ʀ/	/j/
Allophones	[r]	[r̥]	[l]	[l̥]	[ʀ], [ʀ̥]	[j]
Grapheme	⟨r⟩	⟨ř⟩	⟨l⟩	⟨l̥⟩	⟨ʀ⟩, ⟨ʀ̥⟩ ²⁶	⟨j⟩

The voiced obstruent /v/ is realised as [v] before a vowel or a sonorant consonant. Word-finally, /v/ is pronounced as [u]. Before an obstruent consonant /v/ is pronounced [u] word-internally, but word-initially it is pronounced as [v] before a voiced obstruent and as [f] before a voiceless obstruent consonant.

- (12) Sk. **lev** [lɛv] *lion*, **hlavný** [ˈhɫavniː] *main*, **nahrávka** [ˈnahrɑːʊkɑ] *recording*,
vták [ˈftɑːk] *bird*, **vdova** [ˈvdova] *widow*,

2.5.3. Czech Phonemes

The presentation of the Czech phoneme inventory is mainly based on Palková (1994).

Czech has five short and five long vowel phonemes.

Phoneme	/i/	/e/	/a/	/o/	/u/
Allophones	[i]	[ɛ]	[a]	[ɔ]	[u]
Grapheme	⟨i⟩, ⟨y⟩	⟨e⟩ ²⁷	⟨a⟩	⟨o⟩	⟨u⟩

²⁶ Before /i/, /í/ the four palatal consonants /ḍ/, /t̥/, /ň/, /ʀ/ are not marked orthographically with their normal diacritical sign. They are only signalled by the use of ⟨i⟩, ⟨í⟩. This means that ⟨di⟩ represents /ḍi/, whereas ⟨dy⟩ represents /di/, etc. Furthermore, these four consonants are always written without their diacritical sign before short /e/, which means that normally ⟨de⟩ represents /ḍe/ etc., but occasionally, especially in foreign borrowings, the same orthography also represents /de/.

²⁷ After the palatals /ḍ/, /t̥/, /ň/, the diacritic sign is carried by **e** instead of by the consonant, i.e. **dě**, **tě**, **ně**. The phoneme /j/ occurring between a labial obstruent and /e/

Phoneme	/i/	/é/	/á/	/ó/	/ú/
Allophones	[i:]	[ɛ:]	[a:]	[ɔ:]	[u:]
Grapheme	⟨í⟩, ⟨ý⟩	⟨é⟩	⟨á⟩	⟨ó⟩	⟨ú⟩, ⟨ů⟩

Standard Czech is predominantly considered to have only one diphthong, /ou/, whereas Common Czech has one more.²⁸ However, these diphthongs are best interpreted as combinations of the simple vowels.

Phoneme	/ei/	/ou/
Allophones	[ɛj]	[ɔu]
Grapheme	⟨ej⟩	⟨ou⟩

Czech has the following consonant phonemes.

Phoneme	/p/	/b/	/f/	/v/	/t/	/d/
Allophones	[p]	[b]	[f]	[v]	[t]	[d]
	[b]	[p]	[v]	[f]	[d]	[t]
Grapheme	⟨p⟩	⟨b⟩	⟨f⟩	⟨v⟩	⟨t⟩	⟨d⟩

Phoneme	/s/	/z/	/š/	/ž/	/k/	/g/
Allophones	[s]	[z]	[ʃ]	[ʒ]	[k]	[g]
	[z]	[s]	[ʒ]	[ʃ]	[g]	[k]
Grapheme	⟨s⟩	⟨z⟩	⟨š⟩	⟨ž⟩	⟨k⟩	⟨g⟩

Phoneme	/c/	/č/	/t/	/ď/
Allophones	[tʰ]	[tʃ]	[c]	[ɟ]
	[ɟ]	[tʃ]	[t]	[c]
Grapheme	⟨c⟩	⟨cz⟩	⟨t⟩, ⟨t⟩ ²⁹	⟨ď⟩, ⟨ď⟩ ²⁹

Phoneme	/x/	/h/	/ř/
Allophones	[x]	[h]	[ř]
	[χ]	[x]	[ř]
Grapheme	⟨ch⟩	⟨h⟩	⟨ř⟩

is also written with the same diacritic mark, i.e. **bě** /bje/ etc. Finally, **mě** represents /mĕje/.

²⁸ In a few foreign words /au/ and /eu/ also occur.

²⁹ Before /i/, /i/ the three palatal consonants /ď/, /ř/, /ň/ are not marked orthographically with their normal diacritical sign. They are only signalled by the usage of ⟨i⟩, ⟨í⟩. This means that ⟨di⟩ represents /di/, whereas ⟨dy⟩ represents /di/ etc.

Phoneme	/r/	/l/	/j/	/m/	/n/	/ň/
Allophones	[r]	[l]	[j]	[m] [m̥]	[n] [n̥]	[ɲ]
Grapheme	⟨r⟩	⟨l⟩	⟨j⟩	⟨m⟩	⟨n⟩	⟨ň⟩, ⟨n⟩ ²⁹

2.6. Palatalisation

Polish consonants, except for the already palatal /ś/, /ź/, /ć/ and /dź/, become phonetically palatalised as soon as they are followed by either /i/ or /j/, e.g. **piwo** /pivo/ [ˈpivɔ] *beer*, **Giewont** /gjevont/ [ˈgjevɔnt]. This palatalisation, however, does not have any phonologically distinctive significance. Only the palatal /ɲ/ [ɲ] is a phoneme and may occur word-finally or be followed by a consonant, which the palatalised consonants may not.

In Slovak and Czech, the palatal consonants are phonemes. In these two languages, palatalisation is not an active phonological process.

2.7. Voicing and Devoicing

All three languages exhibit regressive³⁰ assimilation of obstruent consonants.

- (13) Pl. **wódka** /vud-/ka/ [ˈvutka] *vodka*
Sk./Cz. **platba** /plat-/ba/ [ˈpladba] *payment*

When talking at a normal pace, i.e. without pauses between words, this assimilation also affects the realisation of a word-final obstruent if the following word also begins with an obstruent consonant.

- (14) Pl. **wasz brat** /vaš brat/ [vaʒˈbrat] *your brother*
Sk./Cz. **váš byt** /váš bit/ [va:ʒˈbit] *your flat*

All three languages exhibit final devoicing of obstruents before a pause.

- (15) Pl. **ząb** /zow̥b/ [ˈzɔmp] *tooth*
Sk./Cz. **zub** /zub/ [ˈzup] *tooth*

³⁰ Polish and Czech are traditionally considered to exhibit progressive assimilation of obstruents when the voiced Polish /v/ and Czech /ř/ are preceded by a voiceless obstruent. The assimilated form, however, may equally well be considered the underlying form, in which case the orthography has to be regarded as historically conditioned, e.g. Pl. **twój** /tvuj/ or /tfuj/ [tfuj] *yours*. In Government Phonology all assimilations are treated as a shared feature for voice or voicelessness between two adjacent phonemes, often as a result of spreading from a phoneme in a stronger (superordinate) position to a phoneme in a weaker (subordinate) position. Thus in a root, this feature must in principle have already been shared in the lexicon.

Polish³¹ and Czech also exhibit word-final devoicing of obstruents when a following word begins with a vowel or a sonorant consonant.

(16) Cz. **zub má kaz** [ˈzupmaːˈkas] *the tooth has a cavity*

Slovak, however, exhibits final voicing of obstruents in the same context, i.e. word-finally, obstruents are pronounced as voiced if a following word begins with a vowel or a sonorant.

(17) Sk. **zub má kaz** [ˈzubmaːˈkas] *the tooth has a cavity*
Sk. **brat a sestra** [ˈbradaˈsɛstra] *brother and sister*

Assimilation also affects the realisation of prepositions ending in an obstruent consonant.

(18) Pl. **w Krakowie** /v-/ /krakovje/ [ˈfkrɑˈkɔvʲjɛ] *in Cracow*,
Sk./Cz. **v Kodani** /v-/ /kodaři/ [ˈfkɔdɑɲi] *in Copenhagen*

However, with one small exception for Czech, in all three languages the final obstruents of prepositions³² are realised as if word-internal, i.e. the preposition and the following word constitute one phonological unit. Therefore the principles of final (de)voicing are not applied to prepositions.

(19) Pl. **pod mostem** /podmostem/ [pɔdˈmɔstɛm] *under the bridge*

The Czech exception involves the non-syllabic prepositions **k** *towards*, **v** *in*, **s** *with* and **z** *from*. In Standard Czech, words beginning with a vowel are preceded by a voiceless glottal stop. This glottal stop is retained after the four non-syllabic prepositions, but dropped after syllabic prepositions.

(20) Cz. **v autě** /v.aufɛ/³³ [ˈvʲɑuɛɕɛ]³⁴ *in the car*, **pod autem** /podautem/
[ˈpɔdauɛtɛm] *under the car*

³¹ I.e. the prevailing, unmarked variety of Standard Polish. Final voicing, as in Slovak, is however typical for speakers of the marked southern (south-west) regional variety of Standard Polish.

³² Here, prepositions are understood in the narrow sense of the word. Words which are secondarily used as prepositions are not subject to this principle.

³³ There is no reason to include the glottal stop in the inventory of Czech phonemes as it is predictable and without phonological significance. In many varieties of the spoken language it is not pronounced.

³⁴ The assimilation is absolutely regular as the glottal stop is voiceless.

3. Vowel–Zero Alternations

A very large number of the world's languages exhibit vowel–zero alternations to a lesser or greater extent. For instance, French exhibits a certain type of vowel–zero alternation which is mainly of a phonetic character, as it is to a major extent dependent on the pace of speech (Charette 1991). Other languages, e.g. Arabic (Kaye 1990), have vowel–zero alternations which are instead of a phonological character, driven by principles that are regularly triggered in order to adjust co-occurring morphemes to each other.

Slavic vowel–zero alternations are also of this latter phonological type. Many monographs and articles have been devoted to such alternations in different Slavic languages, but the types of alternations on which the main interest has been focused are either stem-internal or prefix-final alternations. Very little attention has been paid to preposition-final vowel–zero alternations.³⁵

Even though stem-internal vowel–zero alternations have been recognised as phonologically driven for quite a long time, it is also significant that in traditional reference grammars vowel–zero alternations have until quite recently been discussed, if at all, mainly when they occur in the morphology (Wróbel 2001, Oravec et al. 1988, Havránek & Jedlička 1981), rather than discussing them coherently within a chapter on (morpho)phonology.

3.1. Historical Background

Today's vowel–zero alternations in Slavic languages are traditionally understood as a consequence of the historical process of elision of weak yers in Late Common Slavic. The two yers in Common Slavic (C.Sl.) were supposedly extremely short vowels: the front vowel /i/ and the back vowel /ǔ/. At the end of the first millennium A.D., the yers developed into full-fledged and stable vowels in strong positions, but in weak positions they were elided. (Bethin 1998: 104ff)

The weak and strong positions can be detected by starting from the end of a word. A word-final yer was weak and therefore elided. A word-internal yer was weak and elided if followed by a syllable containing a stable vowel. Such a stable vowel may either be a vowel other than a yer or the result of a yer in a strong position. Such strong yers were present when they were followed by a syllable containing a yer in a weak position. As expressed in the so-called Havlík's law, counting from the end of a word, every odd-numbered yer was elided in an unbroken sequence of yers, whereas every even-numbered yer

³⁵ Even for a major language like Russian, the subject of vowel–zero alternations in prepositions is only explored to a marginal extent, e.g. by Zotova (2008).

developed into a stable vowel (Havlík 1889, Townsend & Janda 1996: 73f).

- (21) C.Sl. */dīnĭ/³⁶ developed into modern Cz. **den** *day*
C.Sl. */šĭvĭčĭ/ developed into modern Cz. **švec** *shoemaker* (N.sg.)
C.Sl. */šĭvĭca/ developed into modern Cz. **ševce** *shoemaker* (A.G.sg.)

Of course, this is only a summary of the basic facts about the fall of the weak yers. There exists a rich tradition of research on this subject and the developments in the different Slavic languages offer quite a few problems from the point of view of these basic principles, i.e. when the presence or absence of an alternating vowel cannot be explained in a satisfactory way by the historical principles of the fall of weak yers known as Havlík's law. In such cases, one might instead suspect that in one way or another, the data in the modern languages have been reinterpreted by the speakers.

The quality of the vowel which is realised in the position affected by vowel–zero alternation is not the same in all Slavic languages. Because these vowels historically originate from two different yers, one front and one back, some of the Slavic languages exhibit two different vowels to fill in the vocalic position in the context of such alternations, whereas other Slavic languages have unified the system, so that, basically, only one vowel is exhibited in the positions of vowel–zero alternations. In Polish and Czech, the vowel realised is, with extremely few exceptions, /e/, whereas in Slovak both /e/ and /o/ occur.³⁷ This historically explicable duality has later been further enriched by Slovak-specific lengthening and diphthongisation, which has given rise to two more alternants: /ie/ and /uo/. Furthermore, in a few instances, the vowels /i/ and /á/ also occur in positions of vowel–zero alternations in Slovak. However, the exact distribution of the Slovak realisations are not of any significance to the present investigation.³⁸

³⁶ Common Slavic reconstructed forms are mainly based on Old Church Slavonic spelling, where ⟨ъ⟩ is interpreted as representing /i/ and ⟨ѣ⟩ as representing /ü/.

³⁷ The alternation of zero with /i/ (alternatively with /i/ in Slovak and Czech, /y/ in Polish), occurring in imperfective verbs, e.g. Cz. **vybrat** *take out* (pf.) vs. **vybírat** *take out* (imp.), is sometimes also treated as an instance of general vowel–zero alternations. This type of alternation, however, differs significantly from the other, as it is tightly associated with the semantics of verbal imperfective aspect. It is therefore probably better described as an infix used for imperfectivisation of perfective verbs.

³⁸ To my knowledge, no coherent compilation exists of all the alternations occurring in Slovak, only brief mentions with a few examples. (Kráľ & Sabol 1989: 338f., Stanislav 1977: 53, Pauliny 1981: 242). For the most part, these alternations are discussed in connection with the morphological paradigm and/or category in which they occur, e.g. the G.pl. of feminine nouns with a stem ending in a non-palatal consonant (Oravec et al. 1988: 64f.).

3.2. Stem-Internal Alternations

Slavic vowel–zero alternations are most frequent stem-internally. In the majority of cases, a vowel appears in the last syllable of forms which have no inflectional ending, whereas the same vowel is absent in forms which have an inflectional ending. The very same phenomenon may be observed in many other languages as well.

(22)	Swedish	Somali	Slovak
singular	finger <i>finger</i>	jilib <i>knee</i>	orol <i>eagle</i>
plural	fing_r-ar ³⁹	jil_b-o	or_l-y

More complicated cases with more than one alternating vowel-position in the same lexeme are extremely rare. One good example of such multiple alternating vowel positions is found in the Cz. noun **švec** *shoemaker* in (21) above. These types of words were more frequent in older stages of the Slavic languages. Another example is given in (23) below.

(23)	N.sg.	G.sg.	
C.Sl.	*/domůčik-ů/	*/domůčik-a/	<i>house</i> (diminutive)
Old Cz./Sk.	dom_ček	domeč_k-a	
Modern Cz.	domeček	domeč_k-u	
Modern Sk.	dom_ček	dom_ček-a	

Over time, most such forms have been reinterpreted in the process of language change. In this very word, Czech has reduced the number of alternating vowel positions to one. The first of the two historically alternating vowels, which was realised as a vowel in the majority of the declined forms, i.e. in all oblique forms, was reinterpreted as a stable vowel and generalised to the whole paradigm. The Slovak reinterpretation is possibly even more radical, as the old stem of the N.sg. has been adopted as the stem for the whole paradigm, with the effect that the word in modern Slovak no longer has any alternating vowel position at all.

Of course, this is but one quite simple example out of many instances of changes over time that have led to vowel–zero alternations which differ systemically from one Slavic language to another, as well as from the Common Slavic system of alternations based on the fall of weak yers.

³⁹ The underscore is only used as a symbol to illustrate the absence of the alternating vowel, and the hyphen marks the boundary between the stem and the ending. They do not, of course, constitute parts of the languages' standard orthography.

3.3. Prefix-Final Alternations

In prefixes, too, frequent reinterpretations can be observed. As an example, in (24) below, Polish has preserved the historically “correct” vowel–zero alternations, whereas Czech and Slovak have reinterpreted these forms as having a stable vowel in the prefix, leaving a vowel–zero alternation only in the last syllable of the stem.

(24)		<i>he went in</i>	<i>she went in</i>
	C.Sl.	*/vŭ-šidlŭ/ ⁴⁰	*/vŭ-šidla/
	Modern Pl.	wszedł	weszła
	Modern Sk.	vošiel	vošla
	Modern Cz.	vešel	vešla

The prefix in the above Slovak and Czech verbs must probably be considered to consist of allomorphs containing stable vowels, as the form of these prefixes can no longer be explained by the principles which were responsible for the fall of the weak yers.

3.4. Preposition-Final Alternations

Again, when it comes to vowel–zero alternations in prepositions, it is a frequent point of view that a major factor determining the form of such prepositions is still more or less the set of principles determining the fall of weak yers in Common Slavic. Doroszewski (1964: 782f.) argues that the reason for the vocalised form of the preposition **w** *in* in today’s Polish **we Lwowie** *in Lwiv* is a historically-elided yer between the initial consonants of the following noun.

(25)		*/vŭ lŭvo/... ⁴¹
	Pl.	we Lwowie <i>in Lwiv</i>

Supposedly, in (25), the rightmost yer is elided because it is followed by a stable vowel, and the leftmost yer is strengthened into a full vowel because it is followed by a weak yer which is elided.

However, there exists abundant evidence in the modern Slavic languages that such historically-based reasoning is often rather misleading, i.e. that the

⁴⁰ The hyphen is only used to indicate the boundary between the prefix and the rest of the word.

⁴¹ This reconstruction is my own. It should not be interpreted as C.Sl., but purely as an underlying form. It is, however, based on Doroszewski’s (ibid.) reasoning that this prepositional phrase behaves the way it does because the city’s name is based on the same root as the Polish noun N.sg. **lew**, G.sg. **lwa** *lion*, which exhibits vowel–zero alternation in accordance with the behaviour of the yers in C.Sl. */lŭvŭ/ *lion*.

data has been reinterpreted in some way or another, and that different, language-specific principles are at work in each of the Slavic languages.

In order to introduce some data that offer evidence for the probability of such reinterpretations, let us first briefly investigate the already-mentioned prepositional phrase a little further. In (26) we find that the very same preposition behaves differently in two identical phonological environments based on the same nominal root.

- (26) Pl. **w** **Lwowie** *in Lviv*
w lwowskim klubie nocnym *in a Lvovian night club*

Before the noun **Lwowie**, the alternating vowel is realised, whereas it is absent before the corresponding adjective **lwowskim**. According to traditional historical principles of the fall of weak yers, the vocalised form of the preposition would have been expected in both cases, as the common root of both the noun and the adjective, */l'iv/, based on evidence from the C.Sl. noun */l'ivŭ/ *lion*, contains a yer between /l/ and /v/. Consequently, the yer in the preposition, followed by a syllable containing another yer, should be realised as a vowel, but in fact, this is evidently not always the case in modern Polish.

Another example is based on comparison of the three major West Slavic languages. It shows that the results of vowel-zero alternation in these three languages differ radically from each other in the very same phonological contexts.

- | | | | |
|------|------------------|---|--------------------|
| (27) | <i>in school</i> | <i>in a Swedish school</i> | <i>from school</i> |
| Pl. | w szkole | w szwdzkiej szkole | ze szkoły |
| Sk. | v škole | (v)vo švédскеj škole ⁴² | zo školy |
| Cz. | ve škole | ve švédské škole | ze školy |

From the examples in (27), it seems that Polish has the most restrictive usage of vocalised forms of prepositions, whereas the usage of vocalised prepositions is most frequent in Czech. Later, we shall find that statistical evidence supports such a hypothesis. This is also the reason why the present investigation will start with Polish, advance to Slovak, and end by investigating the Czech prepositions.

Hopefully, these few examples have shown that common historically-based principles will not be able to explain the reflexes of vowel-zero alternations in prepositions in the different modern Slavic languages. Instead, it is extremely probable that language-specific principles are at work and that the data have been reinterpreted over time, differently in each language, meaning that we cannot expect to find “traces” of C.Sl. yers still triggering the

⁴² In this collocational type, **vo švédскеj** is used at a rate of ca. 90%, whereas **v švédскеj** is used at a rate of ca. 10%.

vowel–zero alternations in the modern languages.

For reference, all the prepositions exhibiting vowel–zero alternations in the three major West Slavic languages are listed below. In the following text, these prepositions will not be provided with English glosses.

3.4.1. Polish

According to the sources which are discussed in chapter 6.2., Polish exhibits the following vowel–zero alternations in prepositions:

- (28) **bez** ~ **beze** *without*
nad ~ **nade** *over, above*
od ~ **ode** *from*
pod ~ **pode** *under, beneath*
przed ~ **przede** *in front of, before; ago*
przez ~ **przeze** *over, by, through*
spod ~ **spode** *from beneath*
sprzed ~ **sprzede** *from in front of*
w ~ **we** *in*
z ~ **ze** *from, of, out of; with*
znad ~ **znade** *from above*

The historical preposition ***k** ~ **ku** *to, towards* has been structurally reinterpreted in Polish, and the historically alternating vowel has evolved into the non-alternating single form **ku**.

3.4.2. Slovak

According to the sources which are discussed in chapter 7.2., Slovak exhibits the following vowel–zero alternations in prepositions:

- (29) **bez** ~ **bezo** *without*
cez ~ **cezo** *over, by, through*
k ~ **ku** *to, towards*
nad ~ **nado** *over, above*
navzdor ~ **navzdory** *despite, in spite of*
od ~ **odo** *from*
pod ~ **podo** *under, beneath*
pred ~ **predo** *in front of, before, ago*
s ~ **so** *with*
skrz ~ **skrze** *through*
spod ~ **spodo** *from beneath*
sprend ~ **spredo** *from in front of*
v ~ **vo** *in*
z ~ **zo** *from, of, out of*

3.4.3. Czech

According to the sources which are discussed in chapter 8.2., Czech exhibits the following vowel-zero alternations in prepositions:

- (30) **bez** ~ **beze** *without*
dík ~ **díky** *thanks to*
k ~ **ke** ~ **ku** *to, towards*
krom ~ **kromě** *except for, besides*
nad ~ **nade** *over, above*
navzdor ~ **navzdory** *despite, in spite of*
od ~ **ode** *from*
poblíž ~ **poblíže** *close to*
pod ~ **pode** *under, beneath*
před ~ **přede** *in front of, before; ago*
přes ~ **přeše** *over, by, through*
s ~ **se** *with*
skrz ~ **skrže** *through*
v ~ **ve** *in*
z ~ **ze** *from, of, out of*
zpod ~ **zpode** *from beneath*
zpřed ~ **zpřede** *from in front of*

4. Corpora and Method

The aim of this study is to describe the regularities that can be observed in the vowel-zero alternations in prepositions in the modern standard varieties of the three major West Slavic languages, Czech, Polish, and Slovak. The corpora have been constructed so that the main bulk of the texts are non-fictional and not older than twenty years. Only a small quantity of fiction has been included, as writers of fiction are less obliged to adhere to the norms of the standard language. Finally, as the Czech and Slovak languages have had a more turbulent modern history than Polish, some older fiction has also been included for these two languages, beginning at the middle of the 19th century. The fictional and older texts have been included in the corpora in order to try and trace possible variation between genres and over time.

There are significant differences between the three corpora. The Polish corpus is mainly built on parliamentary transcripts, whereas the Slovak and Czech corpora mainly comprise texts from the press. However, this should not be a significant problem, as both these genres well reflect the usage of the codified standard languages. The Czech and Slovak corpora also include older fiction, going back almost 200 years in time, whereas the Polish corpus is almost completely made up of texts from the last 20 years. The reason is that the Polish standard language has been considerably more stable for a longer period of time than the Slovak and Czech languages, which both underwent a quite turbulent evolutionary process in the mid 1800s. A certain need for comparisons over time was therefore felt with regard to Czech and Slovak, but not so much with regard to Polish.

The sizes of the three corpora also differ radically. The Polish corpus comprises 5,877,279 tokens, the Slovak corpus 30,875,191 tokens, and the Czech corpus 57,809,698 tokens. These differences are based on the preliminary findings made at an early stage of preparatory work for this investigation (Nilsson 1996, 1998). It was then established that the situation in Polish is far less complex than the situations in Slovak and Czech, and that the situation in Czech is the most complex of the three. It is therefore natural to work with a larger corpus in the investigation of the languages which exhibit a more complex picture with regard to the issue in question, in order to retrieve as many interesting pieces of data as possible.

A listing of the texts in the three corpora can be found in the bibliography at the end of this book.

4.1. The Polish Corpus

Of the total bulk of 5,877,279 tokens in the Polish corpus, fiction constitutes 8%, European Union regulations 3%, factual texts 5%, press 16% and parliamentary transcripts 68%.

The main reason for these proportions is the comparatively small quantity of Polish texts that are easily accessible free of charge on the internet. The archive of parliamentary transcripts at <www.sejm.gov.pl> is relatively unique in Poland in that it offers large numbers of sizeable files on one single site. Another major source of Polish texts, as well as texts in all the other official languages of the E.U., is the European Union regulations at <ec.europa.eu/enterprise/tris/>, but there each file is considerably smaller.

Unfortunately, other freely accessible Polish texts are most frequently either chunked up into very small files, only offered in formats other than plain text, or scattered in small amounts on a wide variety of different sites, which is very ineffective with regard to systematic collection of texts. Despite such difficulties, a number of freely available fiction and factual texts have been included.

4.2. The Slovak Corpus

The Slovak corpus comprises 30,875,191 tokens. Of the total quantity, fiction constitutes 28%, European Union regulations <1%, factual texts 5% and press 67%.

In both the Slovak and the Czech Republics, a tradition of non-profit open access to texts in large archives is rather well-established. Before the more recent era of databases, most newspapers systematically offered their back-issues in large plain text files. Up until the mid 2000s these archives constituted a vast source of easily accessible texts for concordancing.

Nowadays, similar types of large and systematic archives are being compiled as parts of modern e-libraries. In the Slovak Republic, the main fictional e-library is compiled by the newspaper and publishing house Sme at <zlatyfond.sme.sk>.

4.3. The Czech Corpus

The Czech corpus comprises 57,809,698 tokens. Of the total quantity, fiction constitutes 10%, European Union regulations <1% and press 90%.

Until the early 2000s, the situation concerning large newspaper archives was even more favourable in the Czech than in the Slovak Republic. Therefore, the overwhelming proportion of the Czech corpus is made up of texts from the press.

Presently, Prague's municipality library is rapidly building an open access e-library of Czech as well as translated fiction at <e-knihovna.cz>, offering a large number of texts. A significant part of the corpus texts emerge from these collections.

4.4. Extraction and Analysis of Data

The extraction and analysis of data from the corpora is conducted according to the following procedure.

First, using the WordSmith Tools⁴³ concordance programme, every instance of both the vocalised and the non-vocalised form of the alternating prepositions is extracted from the corpora and stored as a concordance file which lists each and every occurrence of the different forms of each preposition in its context.

253 526	od mostu sledovat doleva. Sadem dojdeme ke Gröbovce - novorenesanční budově,	etro 1999 0104_0316.tx
253 527	, zapíjet vodou, spát na holé zemi, a když ke groši přijdeš, hodit jim ho do kádě, –	ch Novy epochalni vylet
253 528	nosí žid balík za balíkem, my skládali groš ke groši, abysme pole přikupovali, voni už	lais Zapadli vlastenci.tx
253 529	do nesmiřlivosti muž, který jak se říká - má ke Grossovi vztah patriarchální, Miloš Zeman	noviny 1996 0518_123
253 530	zadával. 'Jedte jako blázen,' zvolal, 'nejdřív ke Grossovi a Henkeymu na Regent Street	oyle Příběhy Sherlocka
253 531	dokazuje, že Fi garova svatba má daleko ke grotesce, zato blízko k tragikomedii.	ř fronta 1996 0521_123
253 532	ovšem zbývá pokaždé jen nepatrný krůček ke grotesce a možná ještě neznatelnější k	noviny 1996 0518_123
253 533	ke sledování těchto sportů přistoupíte jako ke grotesce. Pak může jít o docela hezkou	ř fronta 1996 0521_123
253 534	Molierova humoru od grácie přes fraškovitost ke grotesce. Ježto Moliere šil své role na	apek O umeni a kulture

Figure 1. Layout of concordance file.

The text line with the prepositional collocation is preceded by the number of the individual collocation and followed by a reference to the text file in which the collocation occurs. The collocations are sorted alphabetically according to the initial letter of the word following the preposition.

The number of occurrences of the different shapes of the prepositions is calculated with respect to the context provided by the initial letter of the following word. Already at this stage, certain patterns will become evident, as the statistical data are divided into two groups: contexts in which almost exclusively one or the other form of the investigated preposition occurs, and contexts in which both forms of the preposition occur with substantial frequency.

For most contexts in which one of the forms of the prepositions is predominant, a small number of contradictory data will probably be at hand. I will call such small numbers of instances that do not comply with the general tendency *counterexamples*. Counterexamples always have to be very closely investigated manually in order to find out whether they really constitute counterevidence, or whether they rather consist of misleading input data, a problem that will be discussed in more detail below. When all irrelevant data has been eliminated, the contexts in which principally only one or the other form of the preposition occurs can be phonologically analysed and defined.

For contexts in which both forms of the prepositions are relatively frequent, two different strategies have to be applied. Firstly, especially for

⁴³ More details about the programming at <<http://www.lexically.net/wordsmith/>>.

larger amounts of data, a deeper statistical investigation has been carried out, taking into account the second letter of the following context word. Meanwhile, smaller amounts of data have been investigated manually, i.e. read through in order to trace possible patterns that can explain the usage of one or the other form of the preposition.

All the output data from the corpus is of course based on occurrences of the prepositions in their orthographical context, and therefore the proper phonological conclusions have been drawn based on a thorough analysis of these orthographical data. Fortunately however, the Slavic orthographies reflect the phonological structure to quite a high degree. In this respect, the most significant complication is instead the fact that prepositions are often vocalised before numerals written with digits, which cannot with absolute certainty be interpreted as corresponding to phonological strings.

Finally, when all the relevant data have been analysed and the phonological contexts for both forms of the prepositions have been detected and defined, a phonological analysis of all the established contexts is undertaken in order to define the phonological principles and processes responsible for the state of affairs detected in the corpus data.

The earlier-mentioned preliminary investigation (Nilsson 1996, 1998) also established small lists of prepositional collocations containing exceptions that cannot possibly be explained by means of those phonological principles that can be defined based on the bulk of the corpus data. Such deviating collocations will have to be regarded as fixed collocations constituting entries in the lexicon of each of the languages investigated.

4.5. Quality of Input Data

The only major problem with the input data is the vocalised form of the Czech preposition *se with*, which is homographic with the Czech reflexive pronoun *se oneself*. Because of this fact, the Czech preposition *s/se with* has been left out of the general corpus-based investigation. Instead, a smaller, manual investigation has been done, and it seems to be possible to generalise quite a bit based on the findings concerning the homorganic preposition *z/ze from, of*. These two prepositions seem to behave very much in the same way, as, disregarding the voicing difference, they are phonetically identical in every other respect. Hence, presuming that these two prepositions behave in the same way, specific searches for possible data contradicting the observed general tendencies have been carried out in the corpus in order to exclude any counterexamples and to, support the hypothesis that the principles governing the choice of form of the Czech prepositions *z/ze* also hold for *s/se*.

Apart from this case, there are no other problematic cases of homography involving the prepositions investigated within any of the three languages. However, many of the texts in the corpora contain certain phrases from other

languages, especially English. This fact gives rise to a few other minor disturbing factors, e.g. the homography of the Pl. preposition **we** *in* and the English pronoun **we**, or the Slovak preposition **so** *with* and the English adverbial particle **so**. This type of disturbing corpus data might generate some misleading output data, but the quantity is so small that it is easily handled manually whenever such data contradict any general tendencies.

Finally, a small number of misprints and spelling mistakes in the texts may generate misleading output data. One of the two most frequent types is the omission of diacritical signs, which results in e.g. Pl./Cz. **ze** *from, of* instead of Pl. **že**/Cz. **že** *that*. The second type is the frequent omission of hyphens at line breaks in some of the texts, giving e.g. **ve jít** instead of **ve-jít** (**vejít**) *enter*. These types of errors occur only to a limited extent and they can very well be taken care of manually whenever such data contradict any general tendencies.

A further kind of problematic typography, leading to false output data, is when a word is emphasised by spaces between each letter of the word. This leads to instances of single letters which are automatically counted as instances of the non-syllabic prepositions **k**, **v**, **w**, **s**, **z**. This type of error is also taken care of manually whenever such data contradict any general tendencies.

There are also many names with foreign spelling, e.g. ⟨c⟩ used for /k/, ⟨z⟩ used for /c/, etc. Other problem types are represented by Roman digits and various abbreviations.

Of course, not all the data can be checked manually for the above types of mistakes, as the quantity of data is far too large. Therefore, for large statistics that do not contradict any general tendencies, any such misleading data will unfortunately be included. It is only when it comes to smaller figures which clearly contradict a general tendency that each individual instance of the preposition can be checked manually in order to exclude the erroneous data.

4.6. Corpus Terminology

Before proceeding to some preliminary remarks on the corpus data on which the following investigation is based, a few notions from the field of corpus linguistics will be introduced, as they are crucial to the following discussion.

The text mass in a computational corpus consists of signs and intervening empty spaces. It is measured in tokens, which means any string of signs delimited on both sides by a space or a punctuation mark.⁴⁴ Tokens may also be referred to as running words. There may be several identical tokens in a

⁴⁴ In the case of WordSmith Tools version 5, tokens consisting of numbers are separated from tokens consisting of letters of the alphabet. In the statistical data presented in this work, e.g. the size of the respective corpora, the tokens consisting of numbers have been disregarded.

corpus. Another way of measuring the size of a corpus is by counting only unique words, called types or wordforms.

A collocation is a combination of two or more word forms which occur together in a corpus.

An occurrence is an instance of a certain word form or collocation which is present in the searched corpus.

The frequency of a word form or a collocation in a corpus may on the one hand be given as absolute, i.e. the actual number of occurrences of this word form or collocation in the corpus, or on the other hand as relative, i.e. the percentual rate at which a certain form occurs in comparison to another form or certain other forms in the context under investigation. The relative frequency of a vocalised form of a preposition is then understood as the percentual rate at which it occurs in a certain context out of all the occurrences of the preposition in both (or all three of) its forms.

4.7 Notational Conventions

Absolute frequency of the two forms of a preposition is graphically presented in running text by first giving the number of occurrences of the form in the mentioned collocation, followed by a slash and the number of occurrences where the alternate form of the preposition occurs in the same collocation, e.g. Sk. **vo dverách** (53/2) *in the door*. This should be interpreted as 53 occurrences of the collocation **vo dverách**, with the vocalised form of the preposition, compared to two occurrences of the collocation **v dverách**, with the non-vocalised form of the preposition.

The relative frequency or the percentual rate of the vocalised form of the preposition in this example is 53 out of 55, i.e. 96%. As a principle, however, the percentual rate is only calculated when the total number of occurrences of both forms of the preposition in a certain context is at least 100. Also, as a principle, the percentual rate of the relative frequency is presented without decimals.

In search strings the symbol * is used to represent any number of any letters, e.g. Cz. **vše*** would include both the basic wordform **vše** *all, everything* as well as any longer forms beginning with these three letters.

5. General Survey of Data

Before undertaking a detailed investigation of the data concerning each of the languages, some preliminary, purely statistical data will be presented in order to give a general, comparative survey of the frequencies of the different forms of the prepositions investigated in the three languages. The absolute frequencies are found in Table 2 below. The meaning of the prepositions as well as their vocalised shapes can be found in section 3.4.

Polish			Slovak			Czech		
	unvoc.	voc.		unvoc.	voc.		unvoc.	voc.
bez	4,113	8	bez	29,273	144	bez	48,721	1,816
						dík	737	11,860
			k	109,598	11,849	k	250,575	40,482
						krom	776	17,652
nad	6,989	59	nad	23,627	238	nad	42,136	630
			navzdor	23	52	navzdor	83	2,727
od	18,514	96	od	79,490	1,038	od	145,907	2,337
						poblíž	2,013	38
pod	5,493	8	pod	28,478	34	pod	37,168	124
przed	5,018	2,334	pred	50,136	746	před	99,877	952
przez	21,484	237	cez	12,216	8	přes	33,335	120
			s	238,982	45,198	s ⁴⁵	436,214	97,532
			skrz	77	208	skrz	323	926
spod	293	2	spod	1,579	0	zpod	282	0
sprzed	192	0	spred	765	4	zpřed	1	0
w	202,705	2,644	v	767,641	70,536	v	1,264,585	257,338
z	93,833	9,279	z	228,172	46,719	z	428,587	93,660
znad	27	0	znad	3	0			
Total	358,661	14,667	Total	1,570,060	176,774	Total	2,791,320	528,194

Table 2. Number of occurrences in the corpora of non-vocalised and vocalised forms of the prepositions investigated.

It should be noted that the sizes of the three corpora differ substantially, but the overall relative frequencies of these prepositions differ only marginally

⁴⁵ Because of the homony of the Cz. preposition *se* *with* and the pronoun *se* *oneself*, the number of vocalised instances of the preposition *s(e)* is not an absolute number. It is statistically based on a sample of 1000 instances of *se*, in which 87 were instances of the vocalised form of the preposition *s*. The figure 97,532 therefore represents 8.7% of the 1,121,058 instances of the wordform *se* which are present in the Czech corpus.

between the three language corpora.

	Polish	Slovak	Czech
Tokens in corpus ⁴⁶	5,877,975	30,876,937	57,835,305
Prepositions, tokens	373,328	1,746,834	3,319,514
Prepositions, frequency	6.4% ⁴⁷	5.7%	5.7%

Table 3. Absolute and relative frequency of the prepositions in their non-vocalised as well as vocalised forms.

In order to make the survey clearer, the percentual frequencies of the vocalised forms of the prepositions are found in Table 4.

Polish		Slovak		Czech	
beze	0.2%	bezo	0.5%	beze	3.6%
nade	0.8%	nado	1.0%	nade	1.5%
ode	0.5%	odo	1.3%	ode	1.6%
pode	0.1%	podo	0.1%	pode	0.3%
przede	31.7%	predo	1.5%	přede	0.9%
przeze	1.1%	cezo	0.1%	přese	0.4%
spode	0.7%	spodo	0.0%	zpode	0.0%
sprzede	0.0%	spredo	0.5%	zpřede	0.0%
znade	0.0%	znado	0.0%		
		ku	9.8%	ke/ku	13.9%
		so	15.9%	se ⁴⁸	18.3%
we	1.3%	vo	8.4%	ve	16.9%
ze	9.0%	zo	17.0%	ze	17.9%
				díky	94.1%
				kromě	95.8%
		navzdory	69.3%	navzdory	97.0%
				poblíže	1.9%
		skrze	73.0%	skrze	74.1%
Total	3.9%	Total	10.1%	Total	15.9%

Table 4. Frequency of the vocalised form of the prepositions.

In Table 4, the first interesting fact is that the total range of vocalisation of

⁴⁶ Tokens consisting of digits are not included.

⁴⁷ The slightly higher frequency of prepositions in Polish than in Slovak and Czech might be interpreted as an effect of the somewhat more analytical character of Polish.

⁴⁸ See footnote no. 45.

prepositions differs considerably between the three languages. The fact that the alternating prepositions are vocalised in 4% of the cases in Polish, 10% in Slovak and 16% in Czech implies that there exist rather major differences between the principles triggering vocalisation of prepositions in the three languages. The triggering factors in Polish are evidently much more restrictive than those in Slovak and Czech, whereas the triggers in Czech give rise to many more vocalisations than those in Polish and Slovak.

5.1. Types of Prepositions

In Table 4, the prepositions have been rearranged in order to show that there is reason to assume that the alternating prepositions fall into three groups, which behave quite differently with respect to vowel-zero alternations.

5.1.1. Asyllabic Prepositions

The statistical data presented above indicate that the prepositions **k, s, w/v, z** behave differently from the rest of the prepositions. This would imply that the principles regulating the occurrence of the vocalised forms are in some respects different for different prepositions. This makes sense as the prepositions **k, s, w/v, z** consist of a single consonant, which is not enough to constitute a phonetically pronounceable syllable. They are therefore often referred to as asyllabic prepositions. Because of this asyllabicity, they are dependent on the following word, as they must be phonetically realised together with the following word. It is therefore natural that the structure of the following word influences the shape of these less independent prepositions.

5.1.2. Syllabic Prepositions

The rest of the prepositions constitute syllables on their own and are therefore referred to as syllabic. From the statistical facts presented in Table 4, it seems that the principles regulating the vocalisation of the syllabic prepositions differ substantially from the principles regulating the asyllabic prepositions. It is probable that the syllabic prepositions are less phonologically dependent on the following word than the asyllabic prepositions, as the syllabic prepositions are fully pronounceable on their own.

Once again, with the exception of Pl. **przed** and Cz. **poblíž**, the syllabic prepositions fall into two distinct groups. In one group the prepositions are extremely seldom vocalised, and in the other they are vocalised with a very high relative frequency. In the first, seldom vocalised group, we find older prepositions, which are common to all three languages under investigation, as well as to most other Slavic languages. Such old prepositions are generally referred to as primary prepositions.

In the second group, we find almost only Czech prepositions. Furthermore, almost all of these prepositions have evolved from other parts of speech at a relatively late point in time. Such newer prepositions are referred to as secondary prepositions.

6. The Polish Data

As has already been pointed out in the previous chapter, Polish is the West Slavic language in which prepositions are vocalised the least. The vocalised form is applied in only approximately 4% of the collocations containing an alternating preposition. This may be compared to 10% in Slovak and 16% in Czech. Therefore, the factors triggering the vocalisation of prepositions in Polish could be suspected to be less complex than the corresponding factors in the other two languages investigated. This assumption is, as will be reported presently, confirmed by the available literature. Because of the relative simplicity of vowel–zero alternations in Polish prepositions, the investigation will begin with the Polish data.

First, all relevant data are retrieved from the Polish corpus and presented systematically. All instances of the vocalised form of each preposition are extracted from the corpus and the context in which they occur is investigated. Next, the possibility of using the non-vocalised form of the same preposition in the same context is investigated in order to single out the different contexts in which the prepositions are vocalised. Based on these data, it will be possible to make some generalisations about the usage of the two forms of the prepositions. In this phase, some very preliminary hypotheses are formulated about the principles at work.

In the next step, the available literature on the subject will be presented and briefly analysed. The existing literature includes a number of dictionaries and reference grammars, as well as some monographs and articles. The findings in the corpus and the facts presented in the literature are compared and the similarities and differences are discussed.

Finally, the findings based on the corpus data and the literature are summed up.

6.1. The Corpus Data

Preposition	Vocalised		Non-vocalised
bez(e)	8	0.2%	4,113
nad(e)	59	0.8%	6,989
od(e)	96	0.5%	18,514
pod(e)	8	0.1%	5,493
przed(e)	2,334	31.7%	5,018
przez(e)	237	1.1%	21,484
spod(e)	2	0.7%	293
sprzed(e)	0	0.0%	192
znad(e)	0	0.0%	27

w(e)	2,644	1.3%	202,705
z(e)	9,279	9.0%	93,833
Total	14,667	3.9%	358,661

Table 5. Absolute and relative frequency of Polish vocalised prepositions compared to the absolute frequency of their non-vocalised equivalents.

Looking at the general statistical data extracted from the Polish corpus, the most striking fact is how the preposition **przed(e)** stands out from the rest of the syllabic prepositions, with a surprisingly high frequency of vocalisation. Apart from that, the data present a homogeneous picture. The asyllabic prepositions are more often vocalised than the syllabic ones, with the exception of **przed(e)**. However, if the absolute frequency of vocalisation is taken into account, the two asyllabic prepositions exhibit a significantly higher number of instances of the vocalised form than any of the syllabic prepositions.

As the absolute frequency of vocalised forms of the syllabic prepositions is generally lower, the investigation will begin with those prepositions.

6.1.1. Syllabic Prepositions

6.1.1.1. **bez(e)**

beze mnie	8	bez mnie	0
		bez mn* ⁴⁹	0
		bez m*	101

Table 6. Vocalisation of the preposition **bez** in the Polish corpus.

For the vocalised form of the preposition **bez(e)**, only 8 instances of the collocational type **beze mnie** *without me* are present in the corpus. No contrasting occurrences of the type **bez mnie** were found in the corpus, nor are there any other collocations with the non-vocalised form of the preposition followed by the same consonant cluster ⟨mn⟩.

⁴⁹ In corpus search strings, an asterisk is used to denote any letter and any number of letters, hence **mn*** stands for any string (or word) beginning with the two letters **mn**.

6.1.1.2. nad(e)

nade chórem	1	nad chórem	0
		nad chó* / nad chu* ⁵⁰	0
		nad ch*	12
nade czymś	1	nad czymś	7
		nad czym	41
nade głowę	1	nad głowę	1
		nad głowami	7
		nad głową	41
nade królestwem	1	nad król*	0
		nad kr*	33
nade mną	12	nad mną	0
		nad mn*	1
nade wszystko	43	nad wszystko	0
		nad wszystkim	11
		nad wszystkimi	19

Table 7. Vocalisation of the preposition **nad** in the Polish corpus.

The vocalised form of the preposition **nad** occurs in six different collocational types in the corpus investigated. Four of these collocations only occur once each, namely **nade chórem** *above the choir*, **nade czymś** *over something*, **nade głowę** *over the head*, and **nade królestwem** *above the kingdom*. It is also a highly interesting fact that all these four collocations occur in the same text, namely in the novel *Bez pamięci* by Ryszard Dziewulski (2003). In the case of **nade czymś** (1/7)⁵¹ and **nade głowę** (1/1), these collocations contrast with quite a few occurrences of the non-vocalised form of the preposition, followed by different forms of the same lemma. In the case of **nade chórem** and **nade królestwem**, there are no contrasting occurrences present in the corpus, but there are quite a few contrasting occurrences with the non-vocalised form of the preposition in similar phonological contexts, i.e. before ⟨kr⟩ and ⟨ch⟩ respectively. Taking all the different facts into account, it therefore seems improbable that Dziewulski's usage would be representative of standard Polish.

The remaining two collocational types **nade mną** *over me* (12/0) and **nade wszystko** *above all* (43/0) are much more frequent. There are no contrasting

⁵⁰ In Polish, ⟨ó⟩ and ⟨u⟩ are pronounced alike and are therefore phonologically identical.

⁵¹ Of the figures appearing in parentheses after a collocational type, the first one refers to the number of instances of that collocational type, while the figure after the slash refers to the number of instances of the corresponding collocational type with the other form of the same preposition.

occurrences with the non-vocalised form of the preposition followed by these same word forms. On the other hand, there exist some contrasting collocations in which the following word begins with the same strings, ⟨mn⟩ and ⟨wsz⟩, respectively, which makes it less probable that these two collocational types are phonologically conditioned.

It is also worth noting that the meaning of **nade wszystko** is not equivalent to the sum of the meaning of the two words **nad(e)** *over, above* and **wszystko** *all*. Instead, this collocation has the meaning *above all, more than anything else* and is used as an adverbial phrase. The contrasting occurrences of **nad wszystkim** *over all* (sg.) and **nad wszystkimi** *over all* (pl.) have a literal meaning, and most of the occurrences in the plural are followed by a noun phrase, e.g. **nadzór nad wszystkimi instytucjami** *control over all institutions*.

6.1.1.3. od(e)

An uninterpretable instance of **ode 21** is found in a parliamentary transcript. Probably the verb **odegrać** *play* was intended instead of **ode**. It is therefore disregarded as a typing error.

ode bluzgi	1	od bluzgi	0
		od bl*	10
ode bulwy	1	od bulwy	0
		od bu*	22
ode gwiazdki	1	od gwiazdki	0
		od gw*	1
ode mnie	91	od mnie	0
		od mn*	4
ode zemsty	1	od zemsty	0
		od ze*	34

Table 8. Vocalisation of the preposition **od** in the Polish corpus.

Once again, Dziewulski (2003) has produced four unique collocations in his novel: **ode bluzgi**⁵² *by the insult*, **ode bulwy** *by the lump*, **ode gwiazdki** *by the star*, **ode zemsty** *by revenge*. The nouns following the preposition are not among the more frequent lexemes, and no contrasting occurrences of the same nouns with the non-vocalised form of the preposition occur in the corpus. It seems suspicious that all these low frequency collocations are found in one single text and that no such collocations are found in other texts. Furthermore, it can easily be shown that Dziewulski systematically uses a very innovative

⁵² Here, this noun is used in the singular. However, according to ISJP'00 it is only used in the plural.

and somewhat bizarre variety of Polish. As an example, the last of the collocations is found in the sentence **Całkiem'és już ode zemsty zaślepion** *You are already totally blinded by revenge*. Apart from the vocalisation, this sentence contains two quite evident deviations from standard Polish, namely the use of the shortened copula 'és instead of **jesteś** *you are*, and the participle **zaślepion** *blinded* without the standard adjectival ending **-y**. The shortened copula does occur quite frequently in colloquial language, but it is hardly ever written, and if written, it is attached directly to the preceding word without any apostrophe. The drop of the masculine nominative singular adjectival ending **-y** does also occur in standard Polish, but only in the case of a handful of very common adjectives and pronouns.

The only frequent collocational type with the vocalised form of the preposition **od(e)** is **ode mnie** *from me, by me*. There are no contrasting occurrences to this collocational type, but there are some occurrences of the non-vocalised form **od** followed by the same string ⟨mn⟩, which suggests that the vocalised form in the collocation **ode mnie** is not phonologically conditioned.

6.1.1.4. pod(e)

podę Albą	5	pod Albą	1
podę drzwiami	1	pod drzwiami	17
		pod drzwi	5
		pod drzewem	4
podę mną	1	pod mną	0
podę mnie	1	pod mnie	0
		pod mn*	4

Table 9. Vocalisation of the preposition **pod** in the Polish corpus.

For this preposition, Dziewulski (2003) has produced five instances of the collocational type **podę Albą**, but he has himself also produced one instance of the contrasting collocation **pod Albą**. Furthermore, from a phonological point of view, the vocalised form of a preposition is extremely unexpected before a following word beginning with a vowel, and so far this has not occurred with respect to other prepositions in the corpus data.

The single collocation **podę drzwiami** *at the door, by the door* occurs in the novel *Sklepy cynamonowe* by Schulz (1934). It is worth noting that the word following the preposition begins with a cluster of no less than four phonemes, namely /dźvj/, but on the other hand it seems quite unlikely that this vocalisation is phonologically conditioned, as there are many occurrences of contrasting collocations with the non-vocalised form of the preposition followed by the same word form, as well as other word forms beginning with

the strings ⟨drzw⟩ and ⟨drz⟩.

Once again, instances of the vocalised form of the preposition are found in collocations with the first person singular personal pronoun: **pode mną** *under me* (with the instrumental case of the pronoun to denote position) as well as **pode mnie** *under me* (with the accusative case of the pronoun to denote direction). Each collocational type occurs only once, but on the other hand, there are no contrasting occurrences present in the corpus.

6.1.1.5. przed(e)

In the data concerning the vocalised form of the preposition **przed(e)**, a few odd collocations occur in the corpus. The collocation **przede 24** occurs in an ungrammatical sentence. Something seems to be missing between the preposition and the number. Further ungrammatical collocations are **przede wszystkich** (2x) and **przede wszyst** (2x), which are probably typing mistakes instead of **przede wszystkim**. All these occurrences are found in the parliamentary transcripts. An obviously dialectal occurrence, **przede syćkim** (= **przede wszystkim**) *above all*, is found in the newspaper *Tygodnik Podhalański*, which is published in the town of Zakopane in the very south of Poland. These collocations are omitted from the data in the following table.

przede chwilą	1	przed chwilą	411
przede mną	64	przed mną	0
		przed mn*	0
przede wszystkim	2262	przed wszystkim	1
przede wszyskim	1	przed wszystkimi	4
		przed wszz*	24

Table 10. Vocalisation of the preposition **przed** in the Polish corpus.

The single instance of **przede chwilą** *a while ago* is found in Dziewulski (2003). This time, there are 411 contrasting instances with the non-vocalised form of the preposition, which clearly indicates the oddness of Dziewulski's usage.

Once again, only the vocalised form of the preposition is attested before the first person singular personal pronoun in the collocational type **przede mną** *in front of me, before me*.

The collocational type **przede wszystkim** is very frequent, and the only contrasting instance, most likely a writing mistake, is found in the parliamentary transcripts, in the form **przede wszyskim**, lacking a ⟨t⟩.

It seems improbable that the usage of the vocalised form in the collocation **przede wszystkim** is phonologically conditioned, as there are many contrasting occurrences of the non-vocalised form of the preposition before the very same string ⟨wsz⟩. Instead, this frequently used adverbial phrase is most

probably encoded in the lexicon, containing the vocalised form of the preposition as a remnant of an older stage of the language, when the principles regulating the vocalisation of prepositions functioned differently from today.

6.1.1.6. przez(e)

przeze mnie	237	przez mnie	3
		przez mn*	10

Table 11. Vocalisation of the preposition **przez** in the Polish corpus.

About **przeze** there is little to say. Once again, the preposition is vocalised before the first person singular personal pronoun, but this time there exist three contrasting occurrences of the non-vocalised form of the preposition before the very same word form. All three of these instances occur in the parliamentary transcripts. The transcripts have already proven to contain a certain number of typing errors, hence the question remains whether these three occurrences are simply mistakes, or whether they might instead reflect a slightly unstable situation with respect to the vocalisation of this collocational type.

6.1.1.7. spod(e)

spode drogi	1	spod drogi	0
		spod dr*	0
spode łba	1	spod łba	0
		spod łb*	0

Table 12. Vocalisation of the preposition **spod** in the Polish corpus.

The preposition **spod(e)** is not very frequent, and only two instances of the vocalised form occur in the corpus. One instance is rather special in that it constitutes part of a name referring to a farming family: **Naprzeciwno nas jest gospodarstwo Gąsieniców spod Drogi (tak się nazywają)**. *In front of us is the farm of the Gąsienica's from below the Road (that is what they are called)*. The example originates from the newspaper *Tygodnik Podhalański*, published in southern Poland, and it is therefore best interpreted as an instance of local usage, and furthermore as a fixed phrase constituting part of a proper name.

The other collocation literally means *from below the head*, but an older meaning of the noun **łeb** is *forehead* (SJP 1900), and WSPA'06 translates the collocation **patrzeć spod łba** idiomatically as *to scowl, to glower*. The context of the occurrence in the corpus is **spojrzał spod łba**, containing the synonymous verb **spojrzeć**, also meaning *to look*, hence there is reason to assume that this collocation with the vocalised form **spode** constitutes an integrated part of a longer lexicalised expression.

6.1.1.8. *sprzed(e), znad(e)*

No instances of the vocalised prepositions *sprzede* and *znade* were encountered.

6.1.2. Asyllabic Prepositions

As already mentioned, the asyllabic prepositions cannot comfortably be pronounced as independent words. Instead they are pronounced together with the following word. The hypothesis is therefore that the asyllabic prepositions are likely to occur in their vocalised form before phonological strings together with which they would otherwise give rise to phonological structures that are disfavoured by the phonological principles of Polish.

6.1.2.1. *w(e)*

First, all instances of the vocalised preposition **w**e have been extracted from the corpus, sorted and counted with regard to the first grapheme of the following word in the context. All occurrences of the non-vocalised form of the preposition in the corresponding contexts have also been extracted and counted for comparison. During the count of the collocations with the vocalised form of the preposition, it became clear that the raw data, i.e. purely computational data, contains certain occurrences that are not instances of the preposition under investigation, such as the abbreviation **WE** (= **Wspólnota Europejska** *The European Community*), which occurs several hundred times, and the English pronoun *we*, occurring just a few times. The corpus also contains a few occurrences of the preposition **w**e in completely uninterpretable collocations which are evidently due to typing mistakes. These types of irrelevant occurrences have been omitted from the data in the following table.

w e a*	1	0%	w a*	3,123
w e c*	14	0%	w c*	8,435
w e d*	14	0%	w d*	15,526
w e f*	180	15%	w f*	1,034
w e i*	1	0%	w i*	6,099
w e k*	40	0%	w k*	13,536
w e l*	39	1%	w l*	2,670
w e ł*	1	0%	w ł*	448
w e m*	95	1%	w m*	7,215
w e s*	2	0%	w s*	23,040
w e ś*	29	2%	w ś*	1,460
w e t*	7	0%	w t*	24,111
w e w*	1,866	21%	w w*	6,960
w e z*	18	0%	w z*	11,511

Table 13. Vocalisation of the preposition **w** in the Polish corpus.

The largest number of tokens triggering the vocalisation of the preposition **w(e)** do themselves begin with the grapheme ⟨w⟩. This is of course a very reasonable and highly expected context for phonologically conditioned vocalisations, as it means that the preposition vocalises in a context where it is identical to the first phoneme of the following context word. This is of course also true before /v/ and /f/; in the latter case, if not vocalised, the /v/ of the preposition would become voiceless due to assimilatory spreading of voicelessness from the following initial /f/.

However, the situation is not that simple. The vocalisation only takes place in 15% of the occurrences before ⟨f⟩ and 21% of the occurrences before ⟨w⟩. Hence there are many contrasting occurrences with the non-vocalised form of the same preposition before the very same graphemes ⟨w⟩ and ⟨f⟩. We will therefore look more closely at all the data by taking the first two graphemes of the following context word into account.

At this stage, the data relating to the non-vocalised form of the preposition is also corrected by clearing it of irrelevant occurrences. One obvious source is instances which are not occurrences of the preposition **w(e)**. There exist about two dozen abbreviations, such as **w.** for **wiek** *century*, **w/w** or **w.w.** for **wyżej wymieniony** *mentioned above* as well as a few initials which represent first names.

There are also many occurrences which blur the data, even though they are occurrences of the preposition **w(e)**. In quite a few instances the preposition is followed by an abbreviation as in **w WDW**, **w WRJ**, **w WSI**, **w WTO**, **w WWF**, **w ww.** (= **w wyżej wymienionym** *in the above mentioned*) and **w W-wie** (= **w Warszawie** *in Warsaw*). When the abbreviations are read out, either as letters or in their exact wording, the phonological string is of course not to the same as the graphical string.

Some instances occur in a heavily dialectal context. Each of the occurrences **we ftoiej** (= **w której** *in which*), **we Warszawie** *in Warsaw*, **we Watrze** *in Watra*, **we Waksmundzie** *in Waksmund*, **we wędzonce** *in the smoked bacon*, **we wiadrak** (= **w wiadrach** *in the buckets*), **we wiaderku** *in the little bucket*, **we wierchy** *into the high mountains*, **we winkielki** *into angels*, **we wojne** *into war*, **we wojsku** *during military service*, **we workach** *in bags*, and **w ftorym** (= **w którym** *in which*), were found in sentences containing other heavily dialectal traits as well, e.g. ⟨om⟩, ⟨s⟩ instead of ⟨ą⟩, ⟨sz⟩. They were all encountered in the southern newspaper *Tygodnik Podhalański*.

Further, some obvious typing mistakes were encountered: **we święto**, **we worek**, **w wnioskiem**, **w wprowadzając** instead of **we święto** *on the holiday*, **we wtorek** *on Tuesday*, **z wnioskiem**, **wprowadzając** *introducing*.

Finally, one instance of an English name occurs with the preposition investigated: **w White Eagle** *in the White Eagle*. From a Polish point of view, the spelling does not at all render the phonological shape of the word form

following the preposition, as it is realised as /ajt/.

In order to give a clear picture of the distribution of **w(e)** in standard Polish, all the above mentioned occurrences have been omitted from the data in following table.

we fa*	1	w fa*	56
		w fe*	40
		w fi*	217
we fl*	5	w fl*	1
		w fo*	564
we fr	172		
		w fu*	142
		w wa*	148
		w wã*	52
we wc*	58		
we wd*	15		
		w we*	136
		w wę*	36
we wg*	1		
we wi*	6	w wi*	1984
we wł*	378	w wł*	1
we wn*	99		
we wo*	1	w wo*	533
		w ó*	83
we wp*	26	w wp*	1
we wr*	279	w wr*	1
we ws*	895	w ws*	5
we wś*	1		
we wt*	68	w wt*	1
		w wu*	1
		w wy*	2983
we wz*	25	w wz*	1

Table 14. Vocalisation of the preposition **w** before words beginning with ⟨f⟩ or ⟨w⟩.

Judging from the data in Table 14, it seems rather convincing that it is the second phoneme of the following context word that is decisive for the choice of the form of the preposition **w(e)**. The hypothesis is therefore at this point that the non-vocalised form of the preposition is used before a following word beginning with /f/ or /v/ if the next phoneme is a vowel, but that the vocalised form is used if the following phoneme is a consonant.

To make this point as clear as possible, the data from Table 14 is

summarised in Table 15 with regard to the quality of the second grapheme of the context word following the preposition.

we wC*/we fC*	2022	99%	w wC*/w fC*	10
we wV*/we fV*	8	0%	w wV*/w fV*	6975

Table 15. Vocalisation of the preposition **w** before words beginning with ⟨f⟩ or ⟨w⟩ followed by either a vowel or a consonant.

Finally, let us have a closer look at the handful of occurrences that contradict the hypothesis formulated.

The eight unexpected occurrences of the vocalised form are **we fabryce** in *the factory* and **we Wiedniu** in *Vienna* (2x) in *Tygodnik Powszechny*, **we wiejskiej** in *the rural*, **we wierzeniach** in *the beliefs*, **we wioskach** (2x) in *the little villages* in the novel *Zjawisko wspólnoty* by Mikołajewska (1999), and **we wodzie** in *the water* in *Tygodnik Podhalański*.

The ten unexpected occurrences of the vocalised form **w** are **w flegmie** in *the flegma*, **w wspólnym** in *the common*, **w wszczęcie** to *a launching* in *Tygodnik Podhalański*, **w własne** in *his own*, **w wprowadzeniu** in *introducing*, **w Wrocławiu** in *Wrocław*, **w wszystkich** in *all*, **w wtorek** on *Tuesday*, **w wzajemnych** in *the mutual* in the parliamentary transcripts, and **w wspięciu** in *the climbing* in the novel *Dwie damy na pozór Fosse'a i de Laclosa* by Sosińska (s.a.).

Even though these contradictory occurrences are found in texts written in standard Polish, it is interesting to note the quite high representation of the southern Polish newspaper *Tygodnik Podhalański*. It is worth noting that many instances are found in the parliamentary transcripts, which have already proven to contain typing mistakes. It is also interesting that as many as four unexpected vocalisations are found in one and the same novel by Mikołajewska (1999).

Returning for a moment to the four occurrences of the vocalised form **we** before words beginning with ⟨wi⟩, such as **we Wiedniu**, these are very problematic cases. Polish orthography renders the sound [j] by either ⟨j⟩ or ⟨i⟩, depending on the preceding context. The mentioned collocation is pronounced [ve'v'jɛdɲu] and it actually complies very well with the principles formulated above, when considering the pronunciation and not the orthography.

For the remaining contexts, other than before ⟨f⟩ and ⟨w⟩, we will proceed directly to investigate the individual context words found in the remaining collocations occurring with the vocalised form **we**. Here as well, some data proves not relevant to standard Polish. Two instances of **we** before a word beginning with a vowel, **we antryju** in *the entrance* and **we izbie** in *the chamber*, occur in the parliamentary transcripts in sentences which are evidently written in dialect. The same holds for the collocations **we swojom** (= **w swoją** in *one's own*), **we świat** into *the world* and **we świecie** in *the world*, all found in *Tygodnik*

Podhalański. The collocation **we snie** is an obvious typing mistake for **we śnie** *in the sleep*. These instances have been omitted from the data in Table 16, which contains the full list of all collocations occurring with the vocalised form of the preposition **w(e)**, followed by context words which begin with another grapheme than ⟨w⟩ or ⟨f⟩. For each of the context words occurring with the vocalised form of the preposition, any contrasting occurrences with the non-vocalised form before the same contextual word form or the same initial consonant cluster, have also been extracted from the corpus.

we czterech	1	w czterech	31
we czwartek	12	w czwartek	60
we czwórkę	1	w czwórkę	0
		w czwór*	2
		w czw*	76
we dnie	1	w dnie	3
		w dn*	7,306
we drzwiach	1	w drzwiach	37
we dwie	2	w dwie	7
we dwoje	2	w dwoje	0
we dworze	1	w dworze	0
we dwóch	4	w dwóch	144
we dwójkę	3	w dwójkę	0
		w dw	230
we kraju	1	w kraju	416
we krwi	39	w krwi	0
		w krw*	4
we Lwowie	39	w Lwowie	0
		w lw*	2
we łzach	1	w łz*	0
we mgle	6	w mgle	1
we mgłach	1	w mgłach	1
we mnie	82	w mnie	2
		w mnie*	37
we mszy	6	w mszy	4
we śnie	21	w śnie	0
		w śnie*	14
we środe	6	w środe	54
we troje	3	w troje	0
		w tro*	64
we trójkę	1	w trójkę	4
we trzech	1	w trzech	66
we trzy	2	w trzy	17

we zbożach	1	w zbożach	0
		w zb*	91
we znaki	16	w znaki	1
we zwyczaj	1	w zwyczaj	6

Table 16. Vocalisation of the preposition **w** before words not beginning with ⟨f⟩ or ⟨w⟩.

From the data in Table 16, it becomes clear that these occurrences are not primarily conditioned by the phonological principles of Polish. In all instances but **we łzach** *in tears* (1/0), there also exist occurrences of the non-vocalised form of the preposition in the same phonological context, often even before the same word form.

The largest and most interesting group of collocational types with the vocalised form **we** involves different forms of the numerals *two*, *three* and *four*. In the corpus, **we dwoje**, **we dwójkę**, **we troje**, **we czwórke** only occur with the vocalised form of the preposition, whereas **w(e) dwie**, **w(e) dwóch**, **w(e) trójkę**, **w(e) trzech**, **w(e) trzy**, **w(e) cztere** mainly occur with the non-vocalised form. After a closer investigation of these collocations, a semantic difference between the collocations with the vocalised and the non-vocalised form of the preposition becomes rather evident. These numerals are used in the accusative case with the preposition **w(e)** in adverbial expressions denoting a certain number of people doing something together as a group. In this function, the forms **dwóch**, **trzech**, **cztere** express that they are all men, **dwie**, **trzy**, **cztery** that they are all women, and **dwoje**, **troje**, **czworo** that they are of mixed sex, whereas **dwójkę**, **trójkę**, **czwórke** do not reveal any information about their sex. In this type of adverbial phrases the numerals are not followed by any noun. Judging from the data in the corpus, it seems that the vocalisation is only possible in these adverbial constructions, and not possible in instances when the numerals are used in their general attributive function, followed by a noun. Furthermore, with the exception of **w(e) trójkę**, it seems that the usage of the vocalised form is obligatory in these types of adverbial expressions denoting groups of people. This would mean that there is a rather clear difference between expressions with **w** + 2, 3 or 4 in the accusative or the locative, used mainly in attributive functions, and **we** + 2, 3 or 4 in the accusative case, used in an adverbial function with reference to a group of people.

Another group of collocational types is constituted by days of the week, namely **we środe** *on Wednesday* (6/54) and **we czwartek** *on Thursday* (12/60). Even though the vocalised form is rather frequent, the non-vocalised **w środe** and **w czwartek** clearly prevail in the corpus.

The remaining collocational types represent individual cases. High frequency is exhibited by the collocational types **we krwi** *in the blood* (39/0), **we Lwowie** *in Lviv* (39/0) and **we śnie** *in the sleep* (21/0). They also remain

uncontradicted by the corpus data. To this group could also be added **we znaki** (16/1), with only one contrasting occurrence. However, for this collocation, the vocalisation is evidently governed by semantic and phraseological factors. In all 16 instances, the vocalised form of the preposition occurs in the longer phrase **dać/dawać się komuś we znaki** *to make life a misery for somebody*, whereas the single contrasting occurrence **w znaki** is used to express the literal meaning *in signs*.

The collocational type **we mnie** *in me* (82/2) should most probably also be considered to belong to the group of obligatory vocalisations in standard Polish, which would be well in line with the vocalisations of the syllabic prepositions before the first person singular personal pronoun. Both of the two non-vocalised occurrences of **w mnie** are found in the same text, the novel *W imię ojca* by Piskulak (2006).

Now only a very small number of collocations with the vocalised form **we** remain. The collocational types **we mgle** *in the fog* (6/1), **we mgłach** *in the fogs* (1/1), **we mszy** *in the Mass* (6/4) can be suspected to exhibit optional vocalisation in standard Polish. In the case of **we mgle** there is even a possibility that it could be a case of obligatory vocalisation, as the single instance of **w mgle** is found in *Tygodnik Podhalański*, which has already proven to conform less to the standard.

One single instance was found of the collocation **we dnie**, occurring in the longer temporal adverbial phrase **i we dnie, i w nocy** *both day and night*. It contains an obsolete locative singular differing from the modern locative **w dniu**, which is used in other contexts. This collocation is therefore extraordinary in two respects. The three occurrences of **w dnie**, on the other hand, actually involve instances of two different lemmas. **W dnie** is used twice meaning *in the bottom*, involving the locative singular of the lexeme **dno** *bottom*, and once in the phrase **w dnie zwyczajne** *on usual days*, in which **dnie** is the regular accusative plural of the lexeme **dzień** *day*.

Two single collocations, **we dworze** *in the yard* (1/0) and **we zbożach** *in the crops* (1/0), are found in the text *Trans – Gombrowicz* by Piskulak (s.a.). A single occurrence of **we łzach** *in tears* (1/0) is found in *Dżajal obrońca Zajry* by Karcz (s.a.). For these three collocational types, no instances of the non-vocalised form of the preposition are found in the corpus. The question therefore remains as to whether these are reflections of the authors' individual usage, or whether they actually represent instances of optional or even obligatory vocalisations in the standard language.

When it comes to the single instance of **we drzwiach** *in the door* (1/37), as opposed to the many instances of the non-vocalised form of the preposition before the same word form, it is interesting to notice that it occurs in the same text as **pode drzwiami** *at the door, by the door*, namely in the novel *Sklepy cynamonowe* by Schulz (1934). The single instance of **we zwyczajaju** *in the habit* (1/6) occurs in the novel *Tygrys* by Wiśniewski (2001), but the corpus also

contains some instances of the non-vocalised form of the preposition in the same context. It would be worth investigating whether these are only reflections of the authors' individual usage, or whether they should rather be interpreted as optional collocational types in the standard language.

Finally, the single instance of **we kraju** *in the country* is contrasted by 419 occurrences of **w kraju**. It is also found in the parliamentary transcripts, which have already proven to contain a number of typing mistakes. Therefore, this occurrence will not be ascribed any particular significance.

6.1.2.2. z(e)

To begin with, all the occurrences of the vocalised form of the preposition **z(e)** were extracted from the corpus. Just as with **we**, many occurrences were found not to contain the preposition **ze**. A dozen collocations were uninterpretable, a handful of instances contained the abbreviation **ZE PAK SA**, and over 200 instances were typing mistakes or instances of dialectal spelling, where **ze** appeared instead of the conjunction **że** *that*. These occurrences are omitted from the data in Table 17, where the numbers of occurrences of the vocalised preposition **ze** are presented with regard to the initial grapheme of the following word. These figures are contrasted with the number of occurrences of the non-vocalised form of the preposition in the same context.

ze c*	4	0%	z c*	2,823
ze d*	10	0%	z d*	7,076
ze j*	2	0%	z j*	2,002
ze l*	11	1%	z l*	1,335
ze ł*	11	4%	z ł*	251
ze m*	209	8%	z m*	2,730
ze r*	2	0%	z r*	3,510
ze s*	4,841	52%	z s*	4,484
ze ś*	916	99%	z ś*	10
ze t*	15	0%	z t*	16,479
ze w*	1,843	27%	z w*	4,920
ze z*	1,081	25%	z z*	3,164
ze ź*	68		z ź*	0
ze ż*	6	2%	z ż*	360
ze 1*	33	4%	z 1*	902
ze 2*	1	0%	z 2*	997
ze 3*	2	2%	z 3*	122
ze 4*	2		z 4*	51
ze 6*	1		z 6*	70

Table 17. Vocalisation of the preposition z in the Polish corpus.

In Table 17, it becomes quite obvious that the vast majority of the vocalisations occur before the sibilants ⟨s⟩, ⟨ś⟩, ⟨z⟩, ⟨ź⟩ and the non-sibilant ⟨w⟩. Phonologically conditioned vocalisation before sibilants would constitute a nice parallel to the phonologically conditioned distribution of **w(e)** before /w/ and /f/ followed by a consonant, and it seems a reasonable working hypothesis that it would again be the second phoneme of the following word that is decisive.

Phonologically conditioned vocalisation before /v/ seems much less plausible, especially as there is not a single instance of vocalisation before the voiceless counterpart ⟨f⟩. A closer investigation of this context will however hopefully explain the high degree of vocalisation before ⟨w⟩.

The number of instances of the remaining types of collocations are considerably smaller and they probably consist of a few collocational types constituting certain fixed phrases, collocations with an optional use of the vocalised form of the preposition, and instances of individual usage.

First we will take a closer look at the distribution of all occurrences of both forms of the preposition **z(e)** before sibilants, i.e. ⟨s⟩, ⟨ś⟩, ⟨z⟩, ⟨ź⟩, ⟨ż⟩. Sorting the data by the two first graphemes of the following word, a manual control for contradictory occurrences can then be undertaken regarding the non-vocalised form of the preposition.

Various sources of misleading data can be observed. In a handful of occurrences **z** simply stands for the letter **z**, mainly as initials representing Christian names. In quite a few other instances the preposition is followed by an abbreviation as in **z SB**, **z SdPL**, **z SKL**, **z SLD**, **z SMGS**, **z SP**, **z SPD**, **z SPO**, **z SRK**, **z SSOŚ**, **z STM**, **z ZChN**, **z ZGM**, **z ZHR**, **z ZSHT**, **z ZSMP**, **z ZSR**, **z ZSRR**, **z ZST**, **z ZSZ**. When the abbreviations are read out, either as letters or in their full wording, the phonological string is of course not the same as the graphical string.

Further, the following obvious typing mistakes were encountered: **z stanie**, **z zmną krwią**, **z zrobiły**, **z związku z tym** instead of **w stanie** *capable*, **z zimną krwią** *could bloodedly*, **zrobiły** *they made*, **w związku z tym** *in connection with that*.

Finally, there are quite a few occurrences of English and German words in which an initial single consonant phoneme is written with more than one consonant grapheme: **z Schengen** *from/with Schengen*, **z science fiction** *with science fiction*, **z Sheffield** *from Sheffield*, **z Shelleyem** *with Shelley*.

When it comes to the vocalised form **ze** occurring before a word beginning with a sibilant followed by a vowel, some instances occur in a heavily dialectal context. Each of the occurrences **ze siedmym** (= **z siódmym** *with the seventh*), **ze sofki** *from the sofa*, **ze Suche** *from Suche*, **ze synem** *with the son*, **ze zadku** *from the rump*, **ze Zębu** *from Ząb*, **ze ziółek** *of herbs*, **ze Zokopanego** (= **z Zakopanego** *from Zakopane*), **ze zowiasów** (= **z zawiasów** *from the hinges*), **ze zyciem** (= **z życiem** *with life*), **ze zyniackom** (= **z żenieniem** *with marrying*)

were found in sentences with other clearly dialectal traits. They were all encountered in the newspaper *Tygodnik Podhalański*.

Finally, the typing mistake **ze zozumierniem** was found instead of the correct **ze zrozumieniem** *with understanding*.

All the above-mentioned problematic occurrences have been omitted from the data in Table 18, which then ought to give quite a clear picture of the situation with regard to the vocalisation of the preposition **z(e)** before words beginning with one of the sibilants ⟨s⟩, ⟨ś⟩, ⟨z⟩, ⟨ź⟩, ⟨ż⟩.

ze sa*	1	z sa*	2,783
		z sạ*	59
ze sc*	53		
		z sd*	1
		z se*	204
		z sẹ*	19
ze sf*	31		
ze si*	1	z si*	325
ze sj*	1	z sj*	2
ze sk*	311	z sk*	1
ze sl*	2		
ze sl̥*	280	z sl̥*	8
ze sm*	31		
ze sn*	13		
ze so*	423	z so*	167
ze sp*	704		
ze sr*	16		
ze ss*	1		
ze st*	1,950	z st*	5
ze su*	1	z su*	101
ze sw*	598		
ze sy*	1	z sy*	257
ze szC*	449	z szC*	4
ze szV*	2	z szV*	255
ze ś*	916		
ze za*	3	z za*	2,674
ze zb*	137		
ze zd*	140	z zd*	1
		z ze*	228
		z zẹ*	35
ze zg*	53		
ze zi*	3	z zi*	87

ze zj*	22		
ze zl*	3		
ze zł*	140		
ze zm*	173		
ze zn*	118		
ze zo*	28	z zo*	27
ze zr*	55	z zr*	1
ze zs*	2		
ze zu*	3	z zu*	47
ze zw*	189		
		z zy*	42
ze ź*	68		
		z źa*	73
		z źą*	17
		z źe*	13
ze źł*	1		
ze źł*	2		
ze źn*	1		
		z źo*	48
		z źó*	9
		z źu*	3
ze źw*	2		
		z źy*	197

Table 18. Vocalisation of the preposition *z* before words beginning with a sibilant in the Polish corpus.

First, two short comments on the contexts in Table 18 can be made. In the case of initial *(sz)*, the third grapheme has also been taken into account, as *(sz)* in Polish is a digraph representing one phoneme only. In the case of *(ś)* and *(ź)*, no further division is needed, as these graphemes are only used before a following consonant or word-finally, hence, we automatically know that these words must begin with a consonant cluster.

The most important additional fact is that all 423 instances of *ze so** are occurrences of the collocational type *ze soba* with each other, with oneself. The corresponding collocation type with the non-vocalised form before the same word form, i.e. *z soba*, only occurs 65 times in the corpus. The behaviour of the collocational types *ze soba* / *z soba* is rather unexpected. They constitute the majority of instances which do not comply with the hypothesis that the vocalised form *ze* is used before a consonant cluster with an initial sibilant, whereas the non-vocalised form *z* is used elsewhere. Summing up the figures in Table 18, we therefore leave aside these two collocational types. The frequencies of the remaining occurrences, distributed according to the hypo-

thesis, are shown in Table 19.

ze SC*	6,462	99%	z SC*	33
ze SV*	43	1%	z SV*	7605

Table 19. Vocalisation of the preposition *z* before words beginning with a sibilant followed by either a vowel or a consonant.

Out of the 33 occurrences of the non-vocalised form of the preposition followed by a word beginning with a sibilant initial consonant cluster, the only noticeable pattern is that the highest number of such instances involves words beginning with (sł), e.g. **z służbami** *with the services*, which is interesting with regard to the fact that /ł/, pronounced [w], may phonologically be regarded as a semi-vowel. It is also interesting to note that 12 of the instances occurred in the newspaper *Tygodnik Podhalański*, nine in the parliamentary transcripts, three in the newspaper *Tygodnik Powszechny*, three in *Sklepy cynamonowe* by Schulz (1934), two in *Zjawisko wspólnoty* by Mikołajewska (1999), and one in *Moja mała Flo* by Siwek (s.a.). The three somewhat strange occurrences were all found in *Dżajal obrońca Zajry* by Piskulak (s.a.): **z sdureguentem** *with the sdureguent*, **z Sjuzy** *from Sjuza*, **z sjuzańskiego** *from the Sjuzian*.

Of the 43 contradictory occurrences containing the vocalised form of the preposition before a word beginning with an initial consonant followed by a vowel, 28 instances were found in the parliamentary transcripts, six in *Tygodnik Podhalański*, six in *Zjawisko wspólnty* by Mikołajewska (1999), and one each in the novels *Sklepy cynamonowe* by Schulz (1934) and *Bez pamięci* by Dziewulski (2003). The only noticeable pattern is that the majority of instances involve **ze zobowiąza*** (19/20) *with/from commitment* and **ze zorganizowan*** (9/0) *with/from organized*, almost all of which occur in the parliamentary transcripts. This might be interpreted as a reflex of the speakers' pronunciation, in order to prevent the preposition from merging with the perfective aspect prefix *z-* of the verb, as this would make it more difficult to hear which aspect is intended, **zorganizowan*** *having been organised* or **organizowan*** *being organised*.

Apart from the two collocational types discussed above, the rest are almost exclusively single occurrences of a row of different collocational types. In those instances it seems most probable that they reflect hasty and unrevised writing (especially the instances in the parliamentary transcripts), hyper-correction⁵³ (especially the occurrences of **w** in the southern newspaper *Tygodnik Podhalański*), regional influence (especially the occurrences of **we** in the

⁵³ Because of the awareness in southern Poland of the dialectal overuse of the vocalised form, there is a risk of exaggerating avoidance of this form.

southern newspaper *Tygodnik Podhalański*) or purely personal preferences (especially the instances in fiction).

Summing up, the above data seem to neatly confirm the expected influence of the second phoneme of the following word on the vocalisation of the preposition **z(e)** – with one exception: the collocation **ze sobą** (423/65) which occurs in 87% of all the collocations of the preposition **z(e)** with the word form **sobą**. Hence, **ze sobą** must evidently be considered an optional but strongly prevailing collocation encoded in the lexicon of standard Polish.

The next large group of triggering head words begins with ⟨w⟩, and this is more intriguing. The first step will therefore be to look at the second grapheme. A few occurrences such as, for instance, the misspelt **ze więty** instead of **ze święty** *the holy*, as well as collocations with the abbreviations **z WKP**, **z WSI**, **z WZS**, have been omitted from the data in the following table.

ze wk*	1	z wk*	1
ze ws*	610	z ws*	110
ze wś*	1	z wś*	6
ze wz*	1,229	z wz*	34

Table 20. Vocalisation of the preposition **z** before ⟨w⟩ followed by a consonant in the Polish corpus.

The unique and somewhat exceptional collocation of the vocalised form of the preposition with a word beginning with **wk*** is **ze wkładem** *with the deposit* found in the parliamentary transcripts. All the remaining instances are collocations of the vocalised form **ze** with a word beginning with ⟨w⟩ followed by a sibilant. These collocations are summed up in Table 21.

ze wS*	1,840	92%	z wS*	150
--------	-------	-----	-------	-----

Table 21. Vocalisation of the preposition **z** before words beginning with ⟨w⟩ followed by a sibilant.

The collocations with the vocalised form of the preposition followed by words beginning with ⟨w⟩ followed by a sibilant involve many dozens of different lexemes. These lexemes are presented in the following table.

ze wschod*	61		
ze wsi	22		
ze wsią	1	z wsią	2
ze wska*	57	z wska*	1
		z wsobności	1
ze wspa*	18	z wspa*	2
ze wspi*	5	z wspieraniem	1

ze wspo*	23	z wspo*	1
ze wspó*	89	z wspó*	5
ze wstawiennictwa	1		
		z wstąpieniem	2
ze wstęp*	12	z wstęp*	7
ze wstrętem	5		
ze wstrzą*	2		
ze wstyd*	14		
		z wsuwany	1
		z wszczęcia	1
ze wszech	65		
		z wszechobecny*	3
		z wszechprawdziwą	1
ze wszelk*	4	z wszelk*	22
ze wszystk*	227	z wszystk*	60
		z wszytych	1
ze wściekłości	2	z wściekl*	6
		z wzajemny*	7
ze wzburzenia	2		
		z wzdętymi	1
ze względ*	3		
ze względ*	1,156	z względ*	2
ze wzgórz*	4		
		z wzięcia	1
ze wzmiank*	2		
ze wzmo*	3		
ze wznosz*	1	z wznosz*	1
ze wzor*	5	z wzor*	15
ze wzrastając*	1	z wzrastając*	2
ze wzro*	45	z wzro*	3
ze wzrusz*	7	z wzrusz*	1
		z wzywającymi	1

Table 22. Vocalisation of the preposition z before words beginning with (w) followed by a sibilant.

In the following table these collocations are presented with regard to the phonetic character of the initial consonant cluster.

ze [vz]C*	1224	z [vz]C*	10
ze [fs]C*	287	z [fs]C*	19
		z [fj]C*	1

ze [fɛ]C*	2	z [fɛ]C*	6
ze [vz]V*	5	z [vz]V*	23
		z [vz̥]V*	1
		z [fs]V*	2
ze [fɛ]V*	23	z [fɛ]V*	2
ze [fʃ]V*	296	z [fʃ]V*	87

Table 23. Vocalisation of the preposition *z* before words beginning with ⟨w⟩ followed by a sibilant, grouped by phonetic context.

In the data presented in the two tables above, no further obvious patterns can be traced. It therefore seems highly probable that the use of the vocalised form of the preposition in this context is optional, but extremely frequent. It occurs in over 92% of the instances. At the same time, it is highly interesting that sibilants constitute part of the triggering context, and it seems that the similarity between these sibilants and the preposition could be part of the reason for these vocalisations, even if an intervening /v/ is present in the phonological structure. The question therefore remains, what kind of optionally applied phonological principles could be responsible for the vocalisations in this type of context. This issue will be discussed further in Chapter 9.

Apart from this, it is however possible to make a few small comments about individual collocational types. It seems that the vocalised form of the preposition *z(e)* is absolutely predominant in collocations with inflectional and even derivational forms based on the lemmas *wschód east*, *wzgląd respect*, *wstyd shame*. It would therefore be worth investigating further whether the usage of the vocalised form of the preposition *z(e)* is obligatory or only optional, but highly frequent, in collocations with different forms of these three lemmas.

Two other collocational types are also highly interesting, but for other reasons. The vocalisation in the collocational type *ze wsi* *from the countryside, from a village* (22/0) is not contradicted by the corpus data. It is also remarkable because of its stress pattern, as the stress falls on the preposition: *ze wsi*. This happens only in a handful of collocations of preposition and noun or preposition and pronoun, and this information must therefore be encoded in the lexicon of standard Polish. Further, the collocational type *ze wszech* *from all* (65/0) is interesting. It contains the old genitive plural form *wszech*, which only occurs in a few fixed phrases, as it has been replaced in modern standard usage by the form *wszystkich*. Only the two types *ze wszech miar* *by all means* and *ze wszech stron* *from all directions* occur in the corpus. These circumstances indicate that the vocalised form of the preposition is most probably obligatory in the collocations *ze wsi* and *ze wszech*.

Finally, let us proceed to the few remaining instances of the vocalised form

of the preposition **z(e)** which are not followed by a word beginning with a sibilant or a ⟨w⟩. Also in this portion of data, there are a few instances of collocations which blur the picture. The collocations **ze Lródeł**, **ze migłowca**, **ze rodka**, **ze równoważonym** contain obvious typing mistakes and should correctly be written **ze źródeł** *from the sources*, **ze śmigłowca** *from the helicopter*, **ze środka** *from inside*, **ze zrównoważonym** *with a balanced*. The collocation **ze mnom** (= **ze mną** with me) occurs in a heavily dialectal portion of text, whereas **ze Jeanem** involves a French name in which the letter ⟨j⟩ corresponds to the Polish phoneme /ż/. Such instances have been omitted from the data in the following table.

ze chrztu	1	z chrztu	0
		z chrz*	7
ze czterdziestu	1	z czterdzistu	1
ze czterdzieści	2	z czterdzieści	0
		z czt*	44
ze dwa	4	z dwa	1
ze dwadzieścia	2	z dwadzieścia	0
ze dwóch	1	z dwóch	117
ze dwudziestu	1	z dwudziestu	2
ze dwunastu	2	z dwunastu	3
		z dw*	224
ze jej	1	z jej	173
ze Lwowa	8	z Lwowa	0
ze Lwowem	1	z Lwowem	0
		z lw*	0
ze łba	1	z łba	0
		z łb*	0
ze łzami	10	z łzami	0
		z łz*	0
ze młyna	3	z młyna	0
		z mł*	72
ze mną	175	z mną	0
ze mnie	28	z mnie	0
		z mn*	30
ze troje	1	z troje	0
		z tr*	517
ze trzy	10	z trzy	0
ze trzydzieści	2	z trzydzieści	0
		z trz*	123
ze twierdzeniem	2	z twierdzeniem	7
ze 1*	33		

ze 2*	1	
ze 3*	2	
ze 4*	2	
ze 6*	1	

Table 24. Vocalisation of the preposition **z** in remaining contexts.

With the exception of the collocations with words beginning with ⟨lw⟩, ⟨lb⟩ and ⟨łz⟩, it is clear that these vocalisations cannot be conditioned by the phonological strings at the beginnings of words, as the non-vocalised form of the preposition occurs in the same contexts. On the other hand, the number of words in Polish which begin with the three strings mentioned is very small.

The most striking category of collocations with the vocalised form of the preposition are collocations with numerals. Many of the collocational types do not exhibit any contrasting occurrences (or just one) with the non-vocalised form of the preposition, i.e. **ze czterdzieści** (2/0), **ze dwa** (4/1), **ze dwadzieścia** (2/0), **ze troje** (1/0), **ze trzy** (2/0), **ze trzydzieści** (2/0). What all these collocations have in common is that the numeral appears in a highly unexpected form, i.e. the common form of the nominative and the accusative. This is odd, as the ablative preposition **z(e)** meaning *from* governs the genitive case, whereas the comitative preposition **z(e)** meaning *with* governs the instrumental case. After a closer look at the sentences in which these collocations occur, it becomes clear that they do not contain any of the two prepositions. Instead, in all of these collocations the form **ze** has an approximative function, meaning *about*, and should not be interpreted as a preposition but as an adverbial particle. This usage obviously has to be considered as a distinct lexeme from the two prepositions. In ISJP'00 this use of **ze** is defined as a prepositive particle.

The single instance of **z dwa** is found in the longer phrase **z dwa razy większym przekonaniem** *with twice as much conviction*, where the numeral only constitutes part of an intervening adverbial complement between the preposition and the noun phrase it governs, hence in this instance no direct grammatical relationship holds between **z** and **dwa**.

Furthermore, in all the remaining cases of collocations with numerals, i.e. **ze czterdziestu** (1/1), **ze dwóch** (1/117), **ze dwudziestu** (1/2), **ze dwunastu** (2/3), there is a direct relationship between **ze**, used as the approximative particle *about*, and **z**, used as preposition in the meanings *from*, *with*.

Concerning the occurrences of **ze** before digits, in the majority of cases, they are occurrences with the lemma **sto** *hundred*, which begins with a sibilant and a following consonant, hence these vocalisations are phonologically conditioned. In a few other instances, the form **ze** is applied in its meaning as approximative particle.

Returning to the rest of the data, the collocational types **ze Lwowa** *from*

Lviv (8/0), **ze mną** *with me* (175/0), **ze mnie** *from me* (28/0) do not exhibit any contrasting collocations with the non-vocalised form of the preposition followed by the same word forms. Also, the number of occurrences is large enough to indicate that the vocalised form is most probably obligatory in these collocational types. In the case of **ze łzami** *with tears* (10/3), the vocalised form is optional but more frequent than the non-vocalised form. Finally, for the collocational types **ze chrztu** *from the baptism* (1/0), **ze Lwowem** *with Lviv* (1/0), **ze łba** *from the head* (1/0), **ze młyna** *from the mill* (3/0), the occurrences are so few that the question still remains as to whether these vocalisations constitute part of standard Polish, and whether they are obligatory or optional, or even serve to express a semantic distinction.

Finally, for **ze jej** *with her* (1/173) and **ze twierdzeniem** *with the claim* (2/7), there are many contrasting collocations with the non-vocalised preposition. Both of these collocational types occur in the parliamentary transcripts, and it therefore seems possible to interpret them as typing mistakes.

6.2. The Literature

We will now proceed to investigate the existing literature on the subject. A brief account will be given of relevant data as well as relevant analyses of these data.

For Polish it is interesting to note that among the linguistic publications that deal with this issue, it is mainly the dictionaries and the language correctness handbooks that are concerned with vowel-zero alternations in prepositions. Grammars and monographs on phonetics and phonology are less so. Therefore, a considerable number of different dictionaries and language correctness handbooks have been included in this investigation.

In the following, I will briefly characterise each of the works and the data put forward will be compared to the earlier-presented corpus-based findings.

6.2.1. Dictionaries

All major monolingual dictionaries have notes on the usage of the vocalised forms of the prepositions. Some of the dictionaries investigated are orthographical, *Słownik ortograficzny języka polskiego* (SOJP'86)⁵⁴ and *Słownik ortograficzny i prawidła pisowni polskiej* (SOPP'90), as these dictionaries list a vast number of collocations including vocalised prepositions, thereby actually implying that they are interpreted as lexicalisations. The language correctness dictionaries *Słownik poprawnej polszczyzny* (SPP'80) and *Nowy słownik poprawnej polszczyzny* (NSPP'99) also give extensive advice on the issue. Among the

⁵⁴ Collective works are referred to with abbreviations, whereas works by individual authors are referred to by the authors' last names.

descriptive dictionaries, five new dictionaries of Polish were taken into account, the four volume *Uniwersalny słownik języka polskiego* (USJP'03), the two volume *Inny słownik języka polskiego* (ISJP'00), the one volume *Słownik współczesnego języka polskiego* (SWJP'96), the one volume *Podręczny słownik języka polskiego* (Sobol 1996), and the fifty volume *Praktyczny słownik współczesnej polszczyzny* (PSWP'95). Two somewhat older dictionaries, the ten volume *Słownik języka polskiego* (SJP'58) and the three volume *Słownik języka polskiego* (SJP'78), have also been taken into consideration. However, the very popular *Mały słownik języka polskiego* (MSJP'74) has not been considered here, as it is essentially a short version of SJP'58, and contains almost no information at all on the issue investigated.

The dictionary data can be roughly divided into two types: a) collocations mentioned as having a vocalised preposition; b) spelled out principles for vocalisation of prepositions. In practice, however, these two types are on many occasions not very well distinguished.

The two orthographic dictionaries (SOJP'86, SOPP'90), of course, give only entries with the vocalised forms of prepositions without any further comments, as these dictionaries simply list the forms, and are intended only to give information about spelling. It is interesting, though, that prepositional collocations are included to such a high degree in these dictionaries. This also reflects the fact that many of these collocations are interpreted as lexicalised wholes.

In Table 25 below, a general survey is given of the type of information which is found in the particular dictionaries. Thereafter, the dictionary data are carefully investigated, taking one preposition at a time and comparing it with the findings established on the basis of the corpus data. As will become clear from the following presentation, there are large discrepancies between the information given in different dictionaries. The same is in some cases also true about the different entries within one and the same dictionary.

The dictionary data in the following table are presented according to two parameters, i.e. whether it is pointed out that 1) the vocalised form of the preposition is used in certain fixed phrases (FP) and 2) that the choice of form depends on certain initial consonant clusters (CC) in the following word. Of course, in many cases both criteria are mentioned in the same dictionary for one and the same preposition.

Concerning the fixed phrases, some works state that the form under discussion is used a) *only* in the fixed phrases mentioned (= FPa), i.e. explicitly stating that no other instances occur; b) in the fixed phrases mentioned (= FPb), i.e. not stating whether the full list or mere examples are given; or c) *for example* in the following fixed phrases (= FPc), i.e. explicitly saying that the list is not complete.

Concerning the consonant clusters, the dictionaries either a) define these clusters precisely (= CCa); b) give some non-precise hints about the nature of

these clusters (= CCb); or c) leave the clusters undefined (= CCc).

	SJP '58	SJP '78	SPP '80	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
beze	FPa	FPb	FPa	FPa	FPb	FPa	FPa	FPa
nade	FPb	FPa	FPc	FPa	FPa CCc	FPb	FPa	FPa
ode	FPb	FPc CCc	FPc CCc		FPa	FPb CCc	FPa	FPa
pode	FPb	FPc	FPc	FPa	FPc CCc	FPa	FPa	FPa
przede		FPc CCc	FPc CCc	FPa	FPa CCc	FPa	FPa	FPa
przeze	some inst.	FPa	FPa	FPa		FPa	FPa	FPa
spode			FPc CCc	FPc		FPa	FPa	FPa
sprzede								
znade								
we						FPb CCa	FPc CCa	FPc CCa
ze	FPc CCa	FPc CCa	FPc CCc	FPb CCc	FPa CCc	FPc CCb	FPc CCa	FPc CCa

Table 25. Overview of the data contained in Polish dictionaries

It is interesting, above all, to note the growing preciseness over the years, leading up to the situation in the two newest dictionaries (ISJP'00⁵⁵, USJP'03⁵⁶), which claim that only the two assyllabic prepositions vocalise before certain consonant clusters. They furthermore claim to give the exact context for those vocalisations as well as the complete set of fixed phrases containing the vocalised form of both the assyllabic and the syllabic prepositions.

⁵⁵ The two volume ISJP'00 is based on the PWN publishing house's card files as well as the company's own computational text corpus of approx. 25 million words (ISJP'00: LVI, LVIII).

⁵⁶ USJP'03 is based on SJP'78, ISJP'00 and the PWN publishing house's own text corpus of 66 million words (USJP'03: XXXIV).

6.2.1.1. beze

The unambiguous picture of this preposition in the corpus is confirmed by the dictionaries. All ten dictionaries examined give one single collocation with the vocalised form of this preposition, namely **beze mnie** *without me*. In SJP'58, SPP'80, SWJP'96, NSPP'99, ISJP'00 and USJP'03, it is also emphasised that this is the only collocational type in which **beze** is applied.

6.2.1.2. nade

For **nad(e)**, the data presented in the dictionaries are somewhat richer than the data found in the corpus. The eight collocations listed in the dictionaries are given in the table below.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
nade dnem	†									
nade dniem	sem.	sem.	•	•	•	•		†		†
nade drogą	†									
nade drzwiami		•		(•)		•		†		
nade Lwowem	((•)) ⁵⁷									
nade mną	•	•	•	•	•	•	•	•	•	•
nade mnie	•		•		•			•	•	•
nade wsią		•		(•)		•		†		
nade wszystko	sem.	sem.	•	•	•	•		•	•	sem.

Table 26. Vocalisation of the preposition **nad** according to Polish dictionaries.⁵⁸

WSJP'96, SJP'78, ISJP'00 and USJP'03 claim to give the complete list of

⁵⁷ In the item 'spode', SJP'58 (VIII, 573).

⁵⁸ The meaning of the symbols used in this and following tables:

- vocalisation occurs in the collocation;
- vocalisation occurs only in the collocation(s);
- (•) vocalisation occurs optionally in the collocation;
- ((•)) vocalisation occurs optionally and rarely in the collocation;
- † vocalisation in the collocation is obsolete;
- sem. vocalisation occurs only in an expression that differs semantically from the same collocation with the non-vocalised form of the preposition;
- phr. vocalisation occurs only in a specific longer fixed phrase;
- R the collocation is used in regional speech varieties.

expressions containing the vocalised form **nade**. All ten dictionaries give the collocation **nade mna** *over me, above me* with the instrumental case, but four of them do not mention **nade mnie** with accusative government of the preposition when denoting motion. Furthermore, **nade dniem** *towards daybreak* is given by eight dictionaries, and **nade wszystko** *above all* by nine, whereof two point out that these two collocations are only applied as semantically marked fixed phrases, used adverbially, expressing time and degree respectively. As a natural consequence, this implies that the collocations **nad dniem** *over the day* and **nad wszystko** *over everything* are also possible unmarked forms, i.e. in the literal meaning, even if such a usage is much less frequent. **Nade drzwiami** *above the door* and **nade wsią** *above the village* are mentioned in four dictionaries, of which SOJP'86 adds that the non-vocalised forms are also possible, whereas SWJP'96 and SJP'78 leave the reader in doubt as to whether the usage is obligatory or not. NSPP'99, however, characterises these collocations together with **nade dniem** *towards daybreak* as archaic. No semantic differences between the vocalised and the non-vocalised forms is reported in any of the dictionaries for these three collocations. Finally, SJP'58 gives two more archaic collocations, **nade dnem** *above the bottom* and **nade drogą** *above the road*.

The situation reported by the newest dictionaries, which is based on rich corpus data, is in accordance with the findings in the corpus used by this study.

6.2.1.3. ode

In two of the dictionaries, SPP'80 and SJP'58, quite an extensive list of different collocations with the vocalised preposition **ode** can be found. Almost all these collocations are, however, marked as regional and archaic. NSPP'99 also contains some of these collocations, but marks them as regional, not as archaic. In the otherwise very precise SWJP'96, this preposition in its vocalised form has been omitted. In SPP'80's rich presentation, it is said that **ode złego** is only used in one particular phrase, namely in **zbaw nas ode złego** *deliver us from the evil one* in the Lord's Prayer. The same is also stated by NSPP'99, ISJP'00 and USJP'03. Rather strangely, SJP'78 mentions vocalisation of **od** both before consonant clusters and in fixed phrases, but only gives one single example, namely **ode mnie**.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
ode dna	†/R									
ode drogi	†/R		†/R					R		
ode drzwi	†/R		†/R		(•)			R		
ode dworu	†/R		†/R							

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
ode krwi	†/R									
ode Lwowa			†/R					R		
ode mnie	•	•	•	•	•		•	•	•	•
ode wschodu	†/R									
ode wsi	†/R		†/R		(•)			R		
ode złęgo			phr.		(•)			phr.	phr.	phr.
ode żniwa	†/R		†/R							

Table 27. Vocalisation of the preposition **od** according to Polish dictionaries.

All of the dictionaries, as well as the corpus data, indicate that only the vocalised form is used in the collocation **ode mnie**. Further, there should be no doubt that the vocalised form is compulsory in the reading of the Lord's Prayer, as stated by four of the dictionaries. This collocation is however not present in the corpus, but there is, in accordance with the data in the dictionaries, one instance of the non-vocalised form in a different context, **od złęgo pana** (1/0) *from the evil master*. Quite as expected, however, there is no evidence in the corpus for a synchronic use of any of the remaining collocational types, which are all marked as archaic, regional or infrequent in Table 27 above. Instead, the following instances with the non-vocalised form of the preposition in the corresponding collocational types were encountered in the corpus data: **od dna** (1/0) *from the bottom*, **od drogi** (7/0) *from the road*, **od drzwi** (3/0) *from the door*, **od wschodu** (2/0) *from the east*, **od wsi** (2/0) *from the village*.

6.2.1.4. **pode**

Besides the collocational types encountered in the corpus, the dictionaries list several other collocations, as presented in Table 28 below, but only six out of ten dictionaries mention the collocation **pode mnie** *in under me* (direction) alongside the more common **pode mna** *under me* (position), which is given by all ten. Six dictionaries mention **pode drzwiami** *right outside the door, under the door*, but SJP'78 differs from the rest by not pointing out that this usage is optional. Only three mention the optional vocalisation in the parallel collocation with accusative case, i.e. **pode drzwi**. **Pode Lwów** *outside Lviv* is given in two dictionaries. In SPP'80 this collocation is characterised as archaic/regional, but in SJP'58 just as optional. Only one dictionary mentions **pode drogą** *below the road*.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
pode drogą			•							
pode drzwi	(•)		(•)					((•))		
pode drzwiami	(•)	•	(•)	(•)	(•)			((•))		
pode Lwów	(•)		†/R							
pode mną	•	•	•	•	•	•	•	•	•	•
pode mnie	•		•		•			•	•	•

Table 28. Vocalisation of the preposition **pode** according to Polish dictionaries.

From the above data, just as from the corpus data, it follows that in contemporary Polish the vocalised form of this preposition is only obligatory in the two collocations **pode mną** and **pode mnie** containing the 1st person singular personal pronoun.

As could be expected, there is almost no evidence at all in the corpus for a synchronic use of any of the remaining collocational types mentioned in the dictionaries. Only one single instance of **pode drzwiami** (1/17) occurs, in a novel from 1934, hence the collocational types with the lexeme **drzwi** *door* should probably be regarded not as optional, but rather as archaic, and the descriptions given in ISJP'00 and USJP'03 seem to be the most appropriate from a synchronic point of view.

6.2.1.5. przede

Only eight out of ten dictionaries note **przede mnie** *in front of me* (direction) with the accusative case, while **przede mną** *in front of me* (position) with the instrumental case is mentioned by all of them. Strangely enough, **przede wszystkim** *above all* is not mentioned by Sobol (1996), and only USJP'03 mentions that this collocation constitutes a lexicalised fixed phrase. Five dictionaries list **przede dniem** *before daybreak* without any further information. Only USJP'03 points out that this collocation is archaic. Furthermore, the example given by SJP'58 clearly indicates that this collocation also has the character of a lexicalised adverbial phrase, synonymous with **przed(e) świtem** *before dawn*. Only two dictionaries mention the optional usage of **przede drzwiami** *in front of the door*, and SJP'58 is the only dictionary to mention the rare collocation **przede świtem**.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
przede dniem	•	•		•	•	•				†
przede drzwiami			(•)	(•)						
przede mną	•	•	•	•	•	•	•	•	•	•
przede mnie		•	•	•	•	•		•	•	•
przede świtem	((•))									
przede wszystkim	•	•	•	•	•	•		•	•	sem.

Table 29. Vocalisation of the preposition **przed** according to Polish dictionaries.

The data in the three most modern dictionaries agree well with the findings of our corpus study. As could be expected, there is no evidence in the corpus for a synchronic use of any of the collocational types, which are all marked as archaic or infrequent in Table 29 above. Only instances with the non-vocalised form of the preposition in the corresponding collocational types were encountered in the corpus data: **przed dniem** (71/0), **przed drzwiami** (6/0), **przed świtem** (8/0).

6.2.1.6. **przeze**

Nine out of ten dictionaries agree that in modern Polish **przeze** is only used in the collocation **przeze mnie** *through me*. SJP'58 adds the archaic usage of **przeze drzwi** *through the door*. Sobol (1996) fails to give even basic information.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
przeze drzwi	†									
przeze mnie	•	•	•	•	•	•		•	•	•

Table 30. Vocalisation of the preposition **przez** according to Polish dictionaries.

The dictionaries confirm the findings based on the corpus, where only the non-vocalised form of the preposition occurs in **przez drzwi** (4/0).

6.2.1.7. **spode**

There is agreement between the dictionaries about the vocalisation of **spode** before the 1st person singular personal pronoun. Only Sobol (1996) once again lacks this information.

The collocation **spode łba** is found in all the dictionaries investigated and, according to SPP'80, **spode łba** *stealthily* constitutes a fixed phrase together with the verb **patrzeć** *look*. However, none of the other dictionaries makes the same comment, not even SJP'58 with all its richness of detail. On the contrary, this dictionary gives examples such as **Rzucał na nią spode łba badawcze**

spojrzenie *He stealthily threw searching glances at her*, and **Jak wilk spode łba typie oczyma** *He glowers stealthily like a fox*. It therefore seems more correct to claim that the prepositional collocation constitutes a fixed adverbial phrase which can be used in any context which involves some kind of looking.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
spode Lwowa	((•))									
spode łba	•	•	phr.	•	•	•	•	•	•	•
spode mnie	•	•	•	•	•	•		•	•	•

Table 31. Vocalisation of the preposition **spod** according to Polish dictionaries.

The data in the dictionaries confirm the findings in the corpus. The rare collocation **spode Lwowa** *from outside Lviv*, mentioned by SJP'58, does not occur in the corpus with any of the forms of the preposition. A search in NKJP⁵⁹ for **spod(e) Lwowa** (5/20), as well as a Google search⁶⁰, where the vocalisation rate was 1%, gives many examples of contemporary use of the vocalised form of the prepositions in this collocation, e.g. on discussion forums, which indicates that it might be somewhat less rare than the dictionaries suggest.

Also, the collocational type **spod łba** with the non-vocalised form of the preposition, still meaning *stealthily*, is not as infrequent on the Internet as one might think. Furthermore, the more rarely occurring literal meaning *from under the head* occurs as **spod łba**, e.g. **wyciek oleju spod łba śruby** *oil leakage from under the head of the screw*.

6.2.1.8. **sprzede**

Sprzede mnie is only mentioned in the orthographic dictionaries. None of the descriptive dictionaries includes this entry.

⁵⁹ All data from NKJP were checked on 7 June 2014.

⁶⁰ All Polish data emerging from Google were retrieved in July 2012 using the search syntax "**search string**" **site:pl**, where the search string is placed within citation marks in order to generate a more precise search and **site:pl** generates a search restricted only to pages with an address ending in <.pl>. The exactness of the searches is, however, somewhat problematic as intervening punctuation marks are disregarded by Google.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
sprzede drzwi				(•)						
sprzede mnie				•	•					

Table 32. Vocalisation of the preposition **sprzed** according to Polish dictionaries.

Our corpus did not exhibit any instances of **sprzede**, but in analogy with the other syllabic prepositions, it seems natural that this preposition will also vocalise before the 1st person singular personal pronoun.

6.2.1.9. znade

None of the dictionaries has an entry *znade*, nor do they comment on this form under the entry *znad*. However, the Internet dictionary *Słownik języka polskiego* at <<http://www.sjp.pl/>> states that this form “is only possible in the traditional collocation ‘znade mnie’.”

The vocalised form of this preposition was not found in our corpus. It could however be assumed that this preposition too is vocalised before the 1st person singular personal pronoun, in analogy with the other non-prefixed and prefixed syllabic prepositions. A Google search also confirms this assumption, with a vocalisation rate of 96% in this context.

6.2.1.10. we

For the preposition **w(e)**, all the dictionaries, although with very different wordings, state that this preposition vocalises in certain phonological contexts. The following information can be found in the cited dictionaries.

According to SPP'80, **w** is vocalised before “words beginning with a group of consonants that are hard to pronounce, especially with the labial consonant *w* in the beginning.” SWJP'96 also talks about “certain consonant clusters, particularly with the consonant *w* at the beginning”. Sobol (1996) is even more imprecise, only referring to “some words beginning with a consonant cluster”. SJP'78 says before “words beginning with a consonant group, especially containing a labial consonant”. SJP'58 gives the most detailed presentation. Vocalisation of **w** is said to take place before “some consonant clusters, especially containing a labial consonant (most often *w* at the beginning of the cluster or at the end of it), more seldom in front of other consonants (the occurrence of *we* varies over time as well as from an environmental and territorial point of view; variations are common, they are sometimes made use of in poetry or for stylisation of the language”. This is followed by examples of the usage, introduced by the contexts, “a) before consonant clusters with a labial consonant at the beginning [...] b) before consonant clusters with the

second consonant being labial [...] c) before consonant clusters not containing any labial consonant." It is also pointed out that **we** is used regionally before a single labial consonant.

The picture is so far not very clear. It has however changed in the three newest dictionaries. NSPP'99 says that **we** is used "together with words beginning with the consonant *w* or *f*, after which follows another consonant [...] regionally (in southern Poland), also before words beginning with the consonant *w* or *f*". ISJP'00 and USJP'03 share an identical definition, and both say that **we** is used "on a regular basis before words beginning with a group of consonants, where the first is *w* or *f*". This confirms the findings of our corpus study.

All the dictionaries before NSPP'99, when talking about vocalisation of the preposition **w(e)** before certain other consonant clusters than those beginning with /*v*/ or /*f*/, give the impression that such vocalisations are also triggered by the initial consonant group of the word following the preposition. A number of collocations are given, but the reader is not informed about the status of these collocations, whether or not they are obligatory, or whether or not there are other collocations with the vocalised form of the preposition before the same consonant clusters. It is not until NSPP'99 that it is clearly stated that all collocations apart from the ones triggered by the above-mentioned principle are exceptions and that the dictionary lists the most important such exceptions. ISJP'00 also gives some statistical information, saying that either the exceptional form is obligatory, more frequent than the non-vocalised form, or less frequent than the non-vocalised form. Unfortunately, USJP'03 and NSPP'99 only make a distinction between obligatory and optional vocalisations. On the other hand, NSPP'99, in contrast with ISJP'00 and USJP'03, makes an attempt to indicate the principles for the usage of the vocalised form before numerals, but unfortunately this is achieved simply by contrasting some individual examples, without stating the general principles. The indications given by this dictionary do however confirm the proposed interpretation of the corpus findings on p. 71.

For natural reasons, i.e. the abundance of phonologically triggered vocalisations, the number of collocations with **we** cited in the dictionaries is much larger than for any of the preceding prepositions. There would be no point in listing them all here. Therefore, the following table only comprises the collocations in which the word following the preposition does not begin with ⟨*w*⟩ or ⟨*f*⟩.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
we dwa	•									
we dwie								sem.		
we dwoje	•									
we dwóch	•			•	•				•	•
we dwójkę	•									
we trójkę			(•)						(•)	•
we trzech	•	•	sem.					sem.	•	•
we czterech			sem.					sem.		
we cztery	•									
we czworo		•								
we czci	•									
we czwartek	•		(•)	(•)	•			(•)	((•))	(•)
we dniu				phr.	•/phr					
we drzwiach	•			(•)						
we dworze	•			•	•					
we krwi	•	•		•	•				(•)	(•)
we Lwowie								•	•	•
we łbach	•									
we łbie		•							•	•
we łzach	•	•							(•)	(•)
we mgle	•	•							(•)	(•)
we mnie			•	•	•	•	•	•	•	•
we mszy									((•))	
we śnie	•	•	•	•	•	•	•		•	•
we środe	•			(•)	•			(•)	((•))	(•)
we środku	•									
we znaki	•			phr.	•				•	phr.

Table 33. Vocalisation of the preposition *w* according to Polish dictionaries.

The examples of the use of the vocalised form before numerals seem rather random in the majority of dictionaries. Only the two language correctness dictionaries (SSP'80, NSSP'99) make any attempt at an explanation. It is also highly interesting that two dictionaries indicate the vocalisation in *w(e) trójkę* *three persons together* as optional, as it confirms the finding made based on the corpus data. It should also be pointed out that not all of the examples of collocations with numerals found in the dictionaries are of the adverbial type denoting groups of people. SJP'58, in particular, gives examples consisting of whole sentences, and therefore it is often not clear whether the vocalisation only occurs in some fixed phrase or not. It can however be suspected that this is the case with regard to *we dwa* and *we cztery*, which only occur in SJP'58,

where they are found in the sentences **Ja to z nią we cztery oczy obgadam** *I will talk to her about it between four eyes* and **furka we dwa konie** *a carriage with two horses*.

Apart from the numerals, the dictionaries mention two other fixed phrases, namely **we dnie i w nocy** and **dać się we znaki**. The earlier dictionaries make no comments on the status of any of the vocalisations listed, whereas the newer dictionaries try to present more detailed information, but there are still some differences between the individual works. For comparison, the corresponding Google data retrieved from the Internet is presented in the following table. As well as the collocations listed in the dictionaries, three other collocations with the vocalised form of the preposition were present in the corpus data: **we mgłach** (1/1) *in the fogs*, **we zbożach** (1/0) *in the crops*, **we zwyczaj** (1/6) *in the habit*. They are therefore included with the Google data. **We trzy** *three persons together*, **we troje** *three persons together* and **we czwórce** *four persons together* have also been added in order to complete the survey of collocations involving numerals.

we dwa	134,000	5%	w dwa	2,600,000
we dwie	214,000	8%	w dwie	2,470,000
we trzy	269,000	11%	w trzy	2,160,000
we cztery	45,800	3%	w cztery	1,460,000
we dwoje	5,860,000	99%	w dwoje	86,700
we troje	93,000	51%	w troje	87,600
we czworo	19,400	45%	w czworo	23,500
we dwóch	279,000	1%	w dwóch	26,800,000
we trzech	105,000	2%	w trzech	6,530,000
we czterech	126,000	3%	w czterech	3,530,000
we dwójkę	311,000	76%	w dwójkę	98,400
we trójkę	228,000	41%	w trójkę	325,000
we czwórce	52,800	29%	w czwórce	129,000
we czci	32,600	46%	w czci	38,000
we czwartek	387,000	2%	w czwartek	15,700,000
we dnie	115,000	24%	w dnie	369,000
we drzwiach	200,000	7%	w drzwiach	2,860,000
we dworze	94,100	25%	w dworze	280,000
we krwi	5,050,000	92%	w krwi	417,000
we Lwowie	1,980,000	93%	w Lwowie	142,000
we łbach	97,400	67%	w łbach	48,700
we łbie	583,000	89%	w łbie	74,100
we łzach	161,000	79%	w łzach	42,800
we mgle	2,230,000	96%	w mgle	86,400
we mgłach	44,800	44%	w mgłach	57,300
we mnie	8,110,000	98%	w mnie	147,000

we mszy	434,000	70%	w mszy	186,000
we śnie	2,660,000	91%	w śnie	346,000
we środe	318,000	1%	w środe	36,700,000
we środku	40,000	0%	w środku	40,300,000
we zbożach	9,110	14%	w zbożach	55,200
we znaki	2,080,000	96%	w znaki	78,600
we zwyczaj	48,300	6%	w zwyczaj	795,000

Table 34. Google data (collected 23 July 2012) regarding marginal collocations with **we** listed in dictionaries.

An inspection of one or two dozen examples of each collocational type in the Google data demonstrates that **we dwa** almost only occurs in the two phrases **we dwa konie** *with two horses*, **we dwa kije** *with two sticks*. **We dwie**, **we trzy** and **we cztery** mainly occur as adverbial complements without a following noun. Sometimes they are also followed by the noun **osoby** *persons*. It is difficult to say anything about the possible occurrence of the non-vocalised form of the preposition in the same adverbial function, as the quantity of data is far too large; however, no such occurrences were found. **We dwóch**, **we trzech**, and **we czterech** are only found in the adverbial function without a following noun, whereas the corresponding collocations with the non-vocalised form are only found with a following noun, hence the vocalisation seems to be obligatory. For **we dwoje**, **we troje**, **we czworo**, **we dwójkę**, **we trójkę** and **we czwórke**, it very soon becomes clear that the vocalised form is only used in the adverbial function, whereas the non-vocalised form is used in that same adverbial function as well as in several other functions, hence the vocalisation is optional. However, the figures indicate that the vocalised form ought to be more frequent, especially for **we dwoje**, **we troje** and **we dwójkę**, but probably also with regard to the remaining three collocational types, considering that the figures in the above table also include the other usages of the non-vocalised form of the preposition.

The situation regarding the rest of the collocations listed in the dictionaries exhibits a high degree of agreement with the findings based on the corpus data as well with the Google data. The obligatoriness of the vocalisation in **we mnie** *in(side) me*, **da(wa)ć się komuś we znaki** *make life a misery for somebody* is confirmed. In the Google data, **we mgle** (96%) is more frequent than **we Lwowie** (93%), but in the corpus data **we Lwowie** (37/0) seems obligatory, whereas there is one instance of **w mgle** (1/6) *in the fog*. It is probably the prescriptive tradition, rather than the actual usage, that leads the dictionaries to present **we Lwowie** as obligatory, whereas **we mgle** is said to be optional. **We śnie** *in the sleep* and **we łbie** *in the head* are also described as obligatory, but the rates of vocalisation according to Google are 91% and 89%. **We krwi** (92%) *in the blood* and **we łzach** (79%) *in tears* are correctly described as optional, but **we czci** (46%) *in the honour*, **we łbach** (67%) *in the heads*, **we mgłach** (44%) *in the*

fogs and **we mszy** (70%) *in mass* are not mentioned in the newer dictionaries, with the exception of **we mszy**, which according to ISJP'00 is rare. On the other hand, ISJP'00 correctly reports the two collocations **we czwartek** (2%) *on Thursday* and **we środę** (1%) *on Wednesday* as being rare, but does not mention the more frequent **we dworze** (25%) *in the courtyard*, **we drzwiach** (6.5%) *in the door* and **we dnie** (24%) *in the daytime*, of which the last occurs exclusively in the longer phrase **we dnie (i) w nocy** *day and night* in the Google data examined. The rate of vocalisation of **we dnie** *in the daytime* in this longer phrase is, according to Google, 96%, i.e. practically obligatory.

6.2.1.11. ze

Concerning **z(e)**, it should first of all be pointed out that according to the dictionary data, Polish clearly has three semantically and syntactically different lexemes all exhibiting the same alternation **z ~ ze**. These lexemes are: 1) the comitative **z(e)** (*together*) *with*, governing the instrumental case, 2) the ablative, elative, and partitive **z(e)** (*out*) *of, (down) from*, governing the genitive case, and 3) the approximative adverbial particle **z(e)** *some*, which does not govern any case.⁶¹ This diversity, however, is not always systematically described under the entry or entries **z** and/or **ze**. Especially when discussing vocalisation of these lexemes, all three lexemes are as a rule treated together.

Concerning the phonologically conditioned vocalisations of **z(e)** we can find the following information in the dictionaries.

The older language correctness dictionary, SPP'80, states that "the preposition in the shape *ze* is usually used before words beginning with a cluster of consonants [...] Based on regional habits, i.e. not generally prevailing, is the usage of the shape *ze* instead of *z* in front of words beginning with the consonants: *s, z, ź, ż*, when they occur before vowels".

The newer language correctness dictionary, NSPP'99, says that **ze** is "used before words beginning with the consonants : *s, z, ś, ź, ż, rz, sz*, after which follows another consonant, as well as before words beginning with several consonants; regionally (in southern Poland) also before words beginning with the consonants *z, s, ź, ś*".

Among the somewhat older descriptive dictionaries, SJP'58 says that **ze** occurs "usually before consonant clusters beginning with *s, z, sz, ś*, more

⁶¹ In the literature this form of **ze** is generally referred to as a particle or as a modulant (SWJP 1996). That it is not a preposition, governing, for instance, the accusative is proven by e.g. **ze czterech panowie** *some four gentlemen*, where it is followed by the nominative. It should however be noted that the approximative **z ~ ze** cannot occur after another preposition, as can, for instance, **blisko** *close to*, **około** *about*, in the same function. For example, in the phrase **na około pięciu miejscach** *in about five places* the modulant **około** may in no way be replaced by **z**.

seldom in other positions”, whereas SJP’78 says that it occurs “usually before consonant clusters beginning with the fricative consonants: s, z, sz, ź, ś, ż, more seldom before other consonant clusters”.

The newer SWJP’96 says that **ze** “occurs only in combination with some words beginning with certain consonant clusters”, whereas Sobol (1996) simply states that it is used “before some words beginning with consonant clusters” and PSWP’95 says that “we use the form *ze* in constructions where the following word begins with certain consonants, e.g. *s, ś, sz, w, m, z*”.

It is only the newest dictionaries that give a stringent explanation. ISJP’00 says that **ze** “is used on a regular basis before words beginning with a group of consonants, where the first is *z, s, ź, sz, ż, ś* [...] Often, but not always, **ze** occurs before the groups *wz-, ws-*”. ISJP’00 also adds that “[t]he form *ze* occurs before the words ‘*dwa*’, ‘*dwanaście*’, ‘*dwadzieścia*’, ‘*dwieście*’, ‘*trzy*’, ‘*trzynaście*’, ‘*trzydzieści*’, ‘*trzysta*’, ‘*cztery*’, ‘*czternąście*’, ‘*czterdzieści*’, ‘*cztery-sta*’ and ‘*sto*’. To use the form *z* in these contexts (with the exception of the position before ‘*sto*’) is also possible, but more rare.” USJP’03, which is partially based on ISJP’00, states only that “the form *ze* occurs: on a regular basis before words beginning with a group of consonants, where the first is *z, s, ź, sz, ż, ś*; [...] often before words beginning with the consonant groups *wz-, ws-*”, and leaves out the valuable information about the vocalisation of the adverbial particle meaning *approximately*.

Just as for **we**, the number of collocations with **ze** cited in the dictionaries is very large. The difference between the syllabic and the asyllabic prepositions is thereby evident. Syllabic prepositions are vocalised in very limited contexts, whereas the vocalisation of the two asyllabic prepositions is triggered phonologically in a vast number of collocations. Again, it is not possible to list all the examples in the dictionaries with **ze**. Therefore, the table only contains collocations with the vocalised form **ze** before words beginning with a consonant other than ⟨s⟩, ⟨ś⟩, ⟨z⟩, ⟨ź⟩, or ⟨ż⟩. The exception is the collocation **ze sobą** with *oneself*, which has been included because it is claimed by three of the dictionaries to be optional, and by one dictionary even to be rare.

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
ze dwa		•			•			•		•/phr
ze trzy		•				phr.				phr.
ze trzydziestu								•		
ze czterdziestkę										•
ze wciąż	•								(•)	(•)

	SJP '58	SJP '78	SPP '80	SOJP '86	SOPP '90	SWJP '96	Sobol 1996	NSPP '99	ISJP '00	USJP '03
ze dniem	phr.									
ze drżenia	•									
ze dwora		phr.								
ze łba	•								•	
ze Lwowa	•							•	•	
ze łzami	•								phr.	phr.
ze mną			•	•	•	•	•	•	•	•
ze mnie			•	•	•	•	•	•	•	•
ze mszą	•									
ze sobą						(•)			((•))	(•)
ze wschodem	•									
ze wsi						•				
ze wsią	•									
ze wstrętem	•									
ze wstydu									•	
ze wszech				phr.	phr.	phr.	phr.		phr.	phr.
ze wszystkim				•	•					
ze względu					•				•	
ze wzruszenia	•									
ze									•	
wzruszeniem										

Table 35. Vocalisation of the preposition *z* according to Polish dictionaries.⁶²

The earlier findings based on the corpus data concerning the approximative adverbial particle agree well with the explanation given by ISJP'00. Therefore, the collocations with numerals found in the dictionaries will be left without further comment. The rather random examples in the dictionaries of collocations with words beginning with (ws) or (wz) will also be discussed no further, as the statements in ISJP'00 and USJP'03 agree well with the previous findings in the corpus data, which indicate that the preposition *z(e)* vocalises at a rate of 92% in these two contexts. According to Google data, not even the reported fixed phrase *ze wszech miar* *in every respect*, mentioned by six out of ten dictionaries, vocalises obligatorily, but merely at a rate of 86%.

Concerning the rest of the collocations cited in the dictionaries, the picture is quite varied. The three remaining fixed phrases mentioned in the dictionaries are *równo ze dniem* *right at daybreak* (SJP'58), *fora ze dwora* *get*

⁶² It should be added that PSWP'95 only mentions the collocations *ze mną* *with me* and *ze wschodu* *from the east*.

out of here (SJP'78), and **ze łzami w oczach** *with tears in the eyes* (USJP'03, ISJP'00). The collocation **fora ze dwora** is also a special case because it rhymes with **fora** (only used in this phrase), as the standard genitive case form would be **dworu**.

Once again, the older dictionaries make no comments on the status of any of the vocalisations listed, whereas the newer dictionaries try to present more detailed information. For comparison, the corresponding data retrieved via Google is presented in the following table. Apart from the collocations listed in the dictionaries, three other collocations with the vocalised form of the preposition were present in the corpus data: **ze chrztu** (1/0) *from the baptism*, **ze Lwowem** (1/0) *with Lviv*, **ze młyną** (3/0) *from the mill*. They are therefore included with the Google data.

ze chrztu	43,900	50%	z chrztu	44,500
ze czcią	107,000	46%	z czcią	124,000
ze dniem	25,600	0%	z dniem	9,580,000
ze drżenia	32	0%	z drżenia	28,600
ze dwora	179,000	88%	z dwora	23,400
ze łba	176,000	73%	z łba	64,900
ze Lwowa	263,000	85%	z Lwowa	47,400
ze Lwowem	84,800	70%	z Lwowem	36,200
ze łzami	632,000	89%	z łzami	80,900
ze młyną	28,800	43%	z młyną	38,800
ze mną	33,400,000	100%	z mną	43,100
ze mnie	8,970,000	98%	z mnie	165,000
ze mszą	34,200	51%	z mszą	32,500
ze sobą	29,300,000	94%	z sobą	1,910,000

Table 36. Google data (collected 23 July 2012) regarding marginal collocations with **ze** mentioned in dictionaries.

The Google data indicate that all the collocational types except **ze dniem** *with the day* and **ze drżenia** *from trembling* have quite a high rate of vocalisation. Even if the context is extended to the alleged fixed phrase **równo ze dniem** *right at daybreak*, the rate of vocalisation according to Google remains as low as 2%. For the collocation **ze dwora** *from the yard*, all of the examples found involve the phrase **fora ze dwora**, whereas **z dwora** occurs as well, but most often in other contexts. This means that the rate of vocalisation in the fixed phrase **fora z(e) dwora** *get out of here* must be considerably higher than 88%.

The vocalisation in the collocations **ze mną** *with me* and **ze mnie** *from me* can without hesitation be considered obligatory, **ze sobą** *with oneself* is extremely frequent, but still optional, whereas **ze łba** *stealthily*, **ze Lwowa** *from Lviv*, **ze Lwowem** *with Lviv* and **ze łzami** *with tears* are somewhat less frequent, but still more frequent than the corresponding collocations with the non-

vocalised form of the preposition. The rate of vocalisation in the longer phrase **z(e) łzami w oczach** *with tears in the eyes* is, according to Google, 91%, which is not significantly higher than in the prepositional phrase alone. Also, the fact that the meaning of the longer phrase is identical to the meaning of its components makes it improbable that there would be anything special about the longer phrase.

The remaining group of collocations, consisting of **ze chrztu** *from the baptism*, **ze czcią** *with honour*, **ze młyna** *from the mill* and **ze mszą** *with the mass*, exhibit a rate of vocalisation close to 50%. All these collocations also have in common that they are less frequent in terms of absolute numbers. This may indicate that vocalisation is more favoured, the higher the absolute frequency of the collocation is. An examination of one or two dozen examples of each collocation also suggested that the collocation **ze młyna** *from the mill* mainly occurs in the phrase **Maryna ze młyna** (8,840/711), for which the rate of vocalisation is 93%. However, the remaining 20,000 occurrences of **ze młyna** exhibit quite heterogeneous contexts. For the other three collocations no such patterns were detected.

6.2.2. Language Correctness

Language correctness handbooks are a popular genre in Poland, and many of the books have longer or shorter notes on the usage of the vocalised forms of the prepositions. The more interesting parts of this literature will be summarised below, but so many statements are imprecise or incorrect that only the more intriguing details will be discussed, whereas obvious misinterpretations will only be pointed out without any discussion.

Gaertner & Passendorfer (1950: 97f.) first give the historical background to the vocalisation of prepositions in a lay, but rather detailed fashion. The impression given is that the historical principles described are still the factors triggering vocalisation in modern Polish, and that the difficulty is only to know exactly where the historical *yers*⁶³ existed. Therefore, 'in practice' it is enough to remember the following principles: 1) Vocalisation "normally occurs where a cluster of consonants makes the pronunciation difficult. [...] 2) Traditionally, the floating *e* is kept in some set combinations of words and some fossilised phrases. [...] 3) We say [...] *nad drzwiami, pod drzwiami*, but sometimes also *nade drzwiami, pode drzwiami*... [...] 4) Say *w Warszawie* [...] etc. (not *we Warszawie* [...]). NB. the floating *e* is generally more common in speech than in writing, and more common in colloquial, spontaneous speech than in careful, scenic speech." (ibid.)

Precise principles are formulated, but some of the statements are incorrect.

⁶³ The notion of *yers* is introduced on p. 40.

It is easy to prove that the historical principles are not responsible for present day vocalisations. An example is *sъ рѣsami > **z psami** *with the dogs*, where the yer of the preposition is followed by a second yer in the following noun, hence the preposition should historically become vocalised, but in modern Polish it is not (Szpyra 1989: 216). Neither is vocalisation necessary before all difficult consonant clusters, as we have, for instance, **z pstrągiem** *with the trout* (Gussmann 2007: 235). The following statements in 2), 3) and 4) are obviously correct, but no principles are formulated, only examples are given. The final NB is highly interesting, but difficult to verify.

In his three volume *O kulturę słowa*, Doroszewski (1964; 1968; 1979) repeatedly discusses the issue. In the first volume he starts by pointing out that the form of the preposition is dependent on "certain historical phonetical conditions" (Doroszewski 1964: 782). A brief traditional background is given and, according to the author, the historical principle for the vocalisation of yers is still a major principle in modern Polish. Besides the historical principle, there are however other principles. The vocalised forms of prepositions are said to occur in the following instances: 1) where it would be vocalised historically, e.g. **we krwi** *in the blood*; 2) where the first syllable of the following word has pleophony in Russian⁶⁴, e.g. **we wrotach** *in the gateway*; 3) before consonant clusters, especially when the following word begins with the same consonant, e.g. **we Wrocławiu** *in Wrocław*. He adds that vocalised forms before a single consonant are regionalisms, e.g. **we wodzie** *in the water*. There is, however, one case where such a usage is standard, namely **ze sobą** *with oneself*, side by side with **z sobą**. He also suggests that that vocalised forms are especially efficient for emphasis.

The problems with this presentation are very much the same as with the previous one. A totally new dimension, however, is the reference made to Russian pleophony. That this principle is not correct is easily confirmed by a list of words where Russian has pleophony, e.g. Polish **w grodzie** *in the castle*, just like Russian **в городе** /v gorodě/ *in the town*. In 3), reference is correctly made to words beginning with the same consonant, but the statement is still not precise enough. The statements made about regional vocalisation before a single consonant and **ze sobą** *with oneself* being the standard form are confirmed by previous data. The comment about emphasis is interesting, but no evidence confirming the statement is given, nor is any such variation indicated by the corpus data.

In the second volume, the form **we** is said to make the preposition more distinct if the following word begins with a consonant cluster where the first

⁶⁴ Slavic pleophony refers to instances of *CVrC and *CVIC, which in some languages have evolved into CVrVC and CVIVC, e.g. *gard- > Russian **город** /gorod/ *town*, as opposed to languages where it has evolved into CrVC and CIVC, e.g. Polish **gród** *castle*.

consonant is identical to the preposition. The same is said about instances like **we wodzie** *in the water* and **ze sokiem** *with juice*, but in Warsaw they sound odd (Doroszewski 1968: 528).

An important point is made here, as the vocalisation of the asyllabic prepositions saves them from disappearing in pronunciation, just as in standard Polish long consonants may only occur before a following vowel, not before a following consonant. Before a single consonant they therefore do not need to vocalise, as they may be distinctly realised in the form of a long consonant, e.g. **w wodzie** [v:ɔdzɛ] *in the water*. However, in the southern and western dialects, long consonants are simplified and pronounced as short (Dejna 1993: 137), hence there is a need in these dialects to vocalise the asyllabic prepositions any time they occur before an identical or similar consonant, with which they would otherwise become assimilated.

In the third volume, Doroszewski (1979: 103–104) introduces somewhat more detailed information regarding the third principle mentioned above. The vocalised form **ze** is said to be used before consonant groups, e.g. **ze względu** *considering*, **ze zmartwieniem** *with worry*. However, if the second element in a consonant group is ⟨ł⟩, then the non-vocalised form of the preposition is used, e.g. **z młodym** *with a young*, **z głośnym** *with a loud*, **z płaczem** *with crying*. The vocalised form is only used in contexts like **ze złością** *with anger*, and the reason for **ze** before **zł-** is that the preposition would otherwise fuse with the following word and disappear in pronunciation. Vocalisation therefore serves to isolate the preposition. Finally, the author mentions the tendency in southern Poland to use **ze** more often than in other areas. It is also said to be a natural tendency in language in general to give a syntactic element a more distinct phonetic form.

The vocalised form **we** is also said to be used before consonant groups which are difficult to pronounce, especially if the first consonant is ⟨w⟩. It is pointed out that **we trzech** and **we czterech** are only used to refer to groups consisting of this number of persons. Finally, it is claimed that **we** is more common in Cracow than in Warsaw (*ibid.*, 288–289).

Again, the principles presented are not sufficiently distinct, but the tendencies and data presented confirm previous observations. The statement that ⟨ł⟩ as the second element of a consonant cluster blocks vocalisation of the preposition **z** is evidently wrong, but the statement about the numerals seems correct and is consistent with previous observations.

Urbańczyk (1966: 488) discusses the avoidance of the vocalised form of prepositions as a result of hypercorrection. Examples of such instances are given, e.g. **w Wrocławiu**. Forms like **we wodzie** are said to be incorrect, although they are becoming more and more common in large parts of Poland (*ibid.*: 455). It is further pointed out that the stress lies on the alternating /e/ of the preposition before forms of the first person singular personal pronoun, which means that such collocations behave phonetically as one word.

The notion of hypercorrection is interesting as there is a considerable awareness of the too frequent use of the vocalised prepositions in dialects and regional usage, which naturally may lead to avoidance of the vocalised form, even in contexts where it is considered correct according to the norm of the standard language. It is also very interesting that collocations with a pronoun and a vocalised preposition are considered to be ‘one word’ because of the atypical stress on the preposition, i.e. on the penultimate syllable of the ‘word’.

Klebanowska et al. (1985: 40-41) make a clear distinction between **w** and **z**, which mainly vocalise before certain consonant clusters, and other prepositions, which only vocalise in **beze mnie, nade mną, nade mnie, pode mną, pode mnie, przeze mnie, przede mną, przede mnie, nade dniem, nade wszystko** and **przede wszystkim**. The preposition **w** is said to vocalise “in combination with nouns where a consonant cluster occurs in the onset, especially if the first of the consonants is *w*”. Apart from that it is also said to vocalise “with the numerals *trzy* and *cztery*, the noun *trójka*, and with certain collective numerals, e.g. *we trzech, we czterech, we trójce* – meaning ‘in a group of three, four people’; *we dwoje, we troje, we czworo* (correctly also *w czworo*)”. Finally, the preposition **z** is said to vocalise “before words beginning with a group of consonants where the first consonant is *s, ś, z, ż, ź*”, as well as “in the collocations *ze mnie, ze mną, ze sobą* (or *z sobą*)”.

This book describes the basic facts in a rather precise fashion, and the indispensable distinction between syllabic and asyllabic prepositions is made explicit, as well as the different types of contexts in which they vocalise. Only the context for the phonologically triggered vocalisation of **w** is not restrictive enough, and the vocalisation of **z** before ⟨wz⟩ and ⟨ws⟩ is not mentioned. However, the remaining statements are well confirmed by previous data, even if not all the details concerning vocalisation with numerals are covered. The set of enumerated non-phonological vocalisations is also too restricted.

Podracki (1993: 34–35) gives a very organised picture of the vocalisation of prepositions. The relevant prepositions (except **sprzed** and **znad**) are enumerated. The vocalisation of prepositions is further said to have historical roots, but today the main reason for vocalisation is to avoid phonetically difficult consonant clusters. It is claimed that there is no way to establish detailed criteria for when the vocalised forms are used. Instead some hints are given. **We** is used before ⟨w⟩ or ⟨f⟩ followed by another consonant, as well as “in some other phonetically difficult collocations, e.g. *we mnie, we Lwowie, we troje*, etc.” There are also some optional cases, e.g. **w(e) czwartek** on *Thursday*, **w(e) środe** on *Wednesday*, **w(e) młynie** in the mill, **w(e) krwi** in the blood, as well as some collocations where vocalisation is rare, e.g. **w(e) śnie** in the sleep, **w(e) mgle** in the fog. The form **ze** is used before consonant clusters when the first consonant is ⟨s⟩, ⟨z⟩, ⟨sz⟩, ⟨ż⟩, ⟨ś⟩ or ⟨ź⟩, as well as “in some fixed phrases, e.g. *ze mną, ze wzrokiem, ze wschodu*, etc.”. There are also some optional vocalisations, e.g. **z(e) lnu** of flax, **z(e) sobą** with oneself, as well as some

collocations where vocalisation is rare, e.g. **ze wzorem** with the pattern. Sometimes one finds hypercorrect forms as **w Lwowie** in *Lviv*, **z studni** from *the well*, which should not be recommended as they sound artificial. In other cases, in southern and western Poland, the vocalised prepositions are used much more often than in the standard variety, e.g. **we Wiedniu** in *Vienna*, **ze Żywca** from *Żywiec*. Such forms before a single consonant are permitted as regionalisms, but today this usage is retreating. The rest of the prepositions only vocalise exceptionally, in a few fixed phrases, e.g. **beze mnie, przeze mnie, nade mną, pode mną, spode łba, przede wszystkim, ode mnie, nade wszystko**, and more seldom, **nade dniem, pode drzwiami, pode dworem**.

The formulation ‘phonetically difficult clusters’ has already been shown to be incorrect, as vocalisation fails to occur in the majority of really complex contexts. The statement that criteria for vocalisation cannot be established is quite appalling, especially when the triggering criteria are then given in a quite detailed fashion. The only rather misleading information is that collocations where **z** is vocalised before ⟨ws⟩ and ⟨wz⟩ are treated as fixed phrases, and that **we śnie** in *the sleep* and **we mgle** in *the fog* are rare. The statement that the regional type of vocalisation is retreating is quite interesting.

Bańko & Krajewska (1994: 342) give a presentation which is very similar to the preceding, but they also mention **sprzede** and **znade**. The presentation also starts by stating that all the mentioned prepositions vocalise on a regular basis before forms of the 1st person sg. personal pronoun. Then the contexts for the phonologically triggered vocalisations are defined, and they are identical to Podracki’s (1993). Vocalisation before single consonants is referred to as regional, and not acceptable in the standard literary language. The following exceptional vocalisations are said to be used “for traditional reasons: *nade wszystko, przede wszystkim, ze wszech miar, we Lwowie, we śnie, ze sobą, ze łzami w oczach, patrzeć spode łba, dać się we znaki, fora ze dwora*, but also in an adverbial function, e.g. *Zjadł ze trzy talerze zupy*.⁶⁵” Finally, examples are given of optional vocalisations: **w(e) trzech, w(e) czwartek, w(e) mgle**.

In this case all the information given is confirmed by the previous data, but more details would be of considerable value, especially concerning the approximative **z(e)** and the collocations with **we** and a numeral.

Kołodziejek (1998: 109-110) also mentions general vocalisation of prepositions before the first person singular personal pronoun, vocalisation of **w** before consonant clusters beginning with ⟨w⟩ or ⟨f⟩, and vocalisation of **z** before clusters beginning with ⟨s⟩, ⟨z⟩, ⟨sz⟩, ⟨ś⟩ or ⟨ź⟩. She also mentions that **z** is often vocalised before ⟨wz⟩ and ⟨ws⟩, and concludes that vocalisation occurs where it would be difficult to pronounce the non-vocalised form of the

⁶⁵ Meaning *He ate some three plates of soup*.

preposition. Furthermore she mentions that regional vocalisation of **w** and **z** before the consonants mentioned also occurs when they are followed by a vowel and she characterises this as sub-standard. Finally, she mentions the following exceptional vocalisations based on tradition: **nade wszystko** *above all*, **przede wszystkim** *above all*, **ze wszech miar** *in every respect*, **we Lwowie** *in Lviv*, **we śnie** *in the sleep*, **ze sobą** *with oneself*, **ze łzami w oczach** *with tears in the eyes*, **patrzeć spode łba** *look stealthily*, **dać się we znaki** *to make life a misery*, **fora ze dwora** *get out of here*, **zjadł ze trzy talerze zupy** *he ate some three plates of soup*, as well as the optional **w(e) trzech** *in three; three persons together*, **w(e) środe** *on Wednesday*, **w(e) mgle** *in the fog*, without any further details.

In this case all the information is also confirmed by the previous data, but unfortunately Kołodziejek fails to mention the vocalisation of **z** before (ż).

Podracki (2000: 114-116) starts by classifying the vocalisations into 1) obligatory, e.g. **we Lwowie** *in Lviv*, **we Wrocławiu** *in Wrocław*; 2) regional, e.g. **we wodzie** *in the water*; 3) optional, e.g. **w(e) czwartek** *on Thursday*; 4) contextual, e.g. **we trzech** *three persons together*, but **w trzech tomach** *in three volumes*. He then gives the historical background to the vocalisation of prepositions, but says that the historical principle does not explain all instances of vocalisation, and that contemporarily the use of the vocalised forms is above all phonetically conditioned, i.e. based on a striving to avoid consonant groups that are difficult to pronounce. Besides that, vocalisations remain in some fixed phrases. He then gives a very detailed list of such fixed phrases, but he does not define in more detail the consonant clusters which are supposedly difficult to pronounce. He lists the following obligatory vocalisations: **beze mnie**, **nade wszystko** *above all*, **nade mną**, **nade mnie**, **ode mnie**, **pode mną**, **pode mnie**, **przede mną**, **przede mnie**, **przede wszystkim** *above all*, **przeze mnie**, **spode mnie**; optional and rare: **pode drzwiami** *right outside the door (position)*, **pode drzwi** *right outside the door (direction)*; contextual: **zbaw mnie ode złego** *deliver me from evil*, **spode łba** *stealthily*; archaic (or regional): **nade dniem** *right before daybreak*, **nade drzwiami** *above the door*, **nade wsią** *above the village*, **ode drzwi** *from the door*, **ode drogi** *from the road*, **ode wsi** *from the village*, **ode Lwowa** *from Lviv*. The forms **we Lwowie** *in Lviv*, **we mnie** *in me*, **ze mną** *with me*, **ze mnie** *from me* are mentioned as being due to the difficult initial consonant cluster, in the same way as the majority of vocalisations of these two prepositions.

Podracki's classification of vocalisation types is very discerning, even though he fails to define the scope of each of the categories. Also, these types only cover one dimension, the obligatoriness of the different vocalisations, whereas the difference between phonologically conditioned and other vocalisations is not clearly pointed out. Quite unfortunately, all instances of the vocalised **we** and **ze** are presented as if they were phonologically triggered.

6.2.3. Grammar

Reference grammars, also including encyclopaediae and textbooks on Polish grammar, are less concerned with the issue under investigation than are the dictionaries and language correctness handbooks. Only rather few of the existing grammars have notes on the usage of the vocalised forms of the prepositions, and the notes are usually rather meagre. Major works such as *Grammatik des Polnischen* (Bartnicka et al. 2004; 620 pp.), *Gramatyka języka polskiego* (Wróbel 2001; 335 pp.), *Nauka o języku dla polonistów* (Dubisz 1999; 623 pp.), *Zarys gramatyki polskiej* (Nagórko 1998; 317 pp.), *Gramatyka języka polskiego dla cudzoziemców* (Kaleta 1995; 487 pp.), and *Gramatyka języka polskiego* (Bąk 1984; 508 pp.) quite surprisingly do not discuss the issue at all.

An early mention of the issue is found in Muczkowski (1849: 123), where it is stated that “*nad, od, pod, przed, w*, and *z* before some nouns beginning with several consonants, as well as before oblique forms of the pronoun *ja*, further *bez* and *przez* only before those forms, take an additional *e*, e.g. *nade drzwiami, nade dniem, pode drzwiami, we Lwowie, ze mną, przede mną, beze mnie* etc.”.

All the relevant prepositions (except **spod, sprzed, znad**) are mentioned, though the information is very general and lacks precision, but the claims are confirmed by the corpus data.

Soerensen (1900: 25) says that the prepositions **bez, nad, od, pod, w** and **z** vocalise according to historical principles “before double consonants”, e.g. **nade mną, nade mnie, ze mną, ze łzami** and **we Lwowie**, but it is only with the two asyllabic prepositions **w** and **z** before unpronounceable double consonants that this principle is consistently followed. Before *muta cum liquida*, the non-vocalised form is used, and ⟨c⟩, ⟨dz⟩, ⟨cz⟩, ⟨dź⟩, ⟨trz⟩, ⟨drz⟩ behave as simple consonants in Polish. He finishes by saying that “firm rules cannot be formulated at all. Generally, a non-vocalised *z* is easier to pronounce than *w*, therefore e.g. *z Lwowa*, but *we Lwowie*”.

This author fails to mention some of the relevant prepositions and most of his statements are imprecise and partly inaccurate. The expression “double consonants” also seems quite misleading. The concept of *muta cum liquida* does not seem to be of any importance for the issue under investigation, as we have, for instance, **we Wrocławiu**, and the statement that **z** would be easier to pronounce than **w** is in no way proven. The frequency of **ze Lwowa** (8/0) in the corpus as well as its relative frequency of 85% in the Google data actually indicates that the vocalised form is the one normally used. The most interesting statement is therefore Soerensen’s list of “simple” consonants.

Meillet & de Willman-Grabowska (1921: 21) say that “[w]hereas we have *w* ‘in’, *z* ‘of’, *od* ‘from’, *nad* ‘above’, *przed* ‘in front of’ before words beginning with a vowel or a single consonant, we have *we, ze, ode, nade, przede* etc. before a group of consonants. [...] Nonetheless we often find the forms without *e* before consonant groups, especially before those where the second element is *r* or *l*.”

These statements are not exactly incorrect, but they are far from precise. Once again reference is made to consonant groups covered by the concept of *muta cum liquida*, which does not seem to be of any importance for this issue.

Łoś (1922: 31) discusses the historical evolution of vowel-zero alternations in prepositions, concluding that the usage of the vocalised forms of prepositions in modern Polish is chaotic, or at least seems to be chaotic, as there has not yet been made any attempt to investigate the issue in a systematic way.

Benni et al. (1923: 116f.) neatly summarise the historical principle saying that the preposition receives an additional *e* when the root syllable of the noun once contained a weak yer, i.e. an ultra-short vowel which was lost, whereas it does not have the additional *e* if the root syllable of the noun has a full vowel. They go on to add that today a new regulation is spreading. It is also pointed out that as a rule in older Polish the preposition was not written at all if the following word began with ⟨s⟩, ⟨z⟩ or ⟨w⟩, e.g. just **wodę** for today's **w wodę** into the water. It is further stated that before nouns in which the root syllable historically had the shape **tort**, **tert**, **tolt** or **telt**, e.g. **środa**, **mleko**, the forms **we** and **ze** were often used in older Polish.

The information is once again correct, but lacks the desired precision with regard to the modern state of affairs. The information about the older spelling is quite interesting, and the final remark reveals the probable reason why Doroszewski (1964: 782), quite incorrectly, claims that Polish prepositions vocalise before words which in Russian exhibit pleophony.

The Czech linguist Kolaja (1947: 150f.), obviously coming from a Czech perspective, says that the prepositions **w** and **z** do not vocalise before ⟨w⟩ and ⟨z⟩ (⟨s⟩), e.g. **w Warszawie**, **z sobą**, but that normally the pronunciation is **we Warszawie**, **ze sobą**. Before other consonants the prepositions are vocalised, e.g. **we środę**, **we Lwowie**.

These statements are quite strange and obviously incorrect. They are probably caused by a striving to prevent Czech learners from transferring certain Czech vocalisations of prepositions into Polish. It is interesting that the pronunciation of certain collocations is said to differ from the way they are written. No other evidence of such differences has been encountered, and it may be suspected that this is a misinterpretation of the difference between the regional and the standard literary norm.

Falkenhahn & Zielke (1957: 195ff.) say that **bez**, **od**, and **przez** vocalise before **mnie**, and that “an *-e* is added to *w* resp. to *z* without any stable rules if a word beginning with several consonants follows”.

Szober (1959: 42) simply states that vowel-zero alternations occur in the prepositions **z(e)**, **w(e)**, **przed(e)**.

Grappin (1963: 31) states that the non-vocalised prepositions are used before a vowel or a single consonant, and the vocalised form before a consonant cluster, except when the second element of the group is a liquid.

All three of these statements are very imprecise.

Damerau (1967: 128f.) says that before a consonant cluster where the first consonant is ⟨w⟩ or ⟨f⟩, the preposition **w** becomes **we**. Likewise, **z** becomes **ze** before consonant clusters beginning with ⟨z⟩, ⟨s⟩ or any other sibilant. He further says that this is originally a reflex of a yer, but that the historical principle is no longer alive. Only a few historically motivated forms still remain in use: e.g. **we mnie** *in me*, **we krwi** *in the blood*, **we Lwowie** *in Lviv*, **ze mnie** *from me*, **ze mną** *with me*, **ze wszystkich sił** *with all one's strength*, **ze wsi** *from the countryside*. Apart from these, some other instances of vocalisation are given simply as examples of the individual prepositions: **beze mnie**, **nade mną**, **ode mnie**, **pode mną**, **przede mną**, **przede wszystkim**, **przeze mnie**.

This statement is principally correct, but lacks a few important details, e.g. the frequent vocalisation of **z** before ⟨wz⟩ and ⟨ws⟩.

Swan (1980: 75) says that "the prepositions *bez, nad, od, pod, przed, przez, w*, **z** appear with a following *e* before certain consonant clusters, and especially before oblique case forms of *ja*". The same statement is also found in Swan (2002a: 42).

Although this statement is principally correct, it is so vague that it almost does not say anything.

A brief note on the issue can also be found in *Encyklopedia języka polskiego* (EJP'91: 272), where it is stated in a short article on prepositions in dialects that "*w* and *z* before the initial consonants *w f z s ś ź sz*, and also *od, nad, pod, przed* before groups beginning with *d-* have in western and southern Poland the vowel *e*: *we wodzie*, [...] *ze solą*, [...] In central, eastern and northern Poland (except for the north-east), the combinations *w wodzie, z solą* prevail".

In this statement the two asyllabic prepositions and the contexts in which they vocalise are mixed together, making the statement inaccurate, or at least unclear. The regional vocalisation of prepositions with a final ⟨d⟩ before words beginning with a cluster containing an initial ⟨d⟩ is also given too much importance, since there is no proof for that statement in the data.

Kucała (1994: 39) also states that "*ze zamku, we Wiedniu, we Wielki Piątek, we wodzie ze sokiem* [...] etc. are forms from the western half of Poland as well as its southern parts."

Bielec (1998: 209) states that in order to "aid pronunciation an **e** is added to the prepositions **bez, nad, od, pod, przed, przez, w**, and **z** before the personal pronouns **mnie** and **mną** and before difficult consonant groups".

The previous two statements are obviously correct, but lack precision.

Kowalik (1999: 112) states that the epenthesis of *e* in asyllabic prepositions prevents the creation of groups of several consonants at morphological borders. The distribution is regulated by phonological factors, and there are also some regional differences. Kowalik also adds that the epenthesis of *e* in syllabic prepositions is as a rule lexically determined. It always takes place before forms of the pronoun **ja** with the stem **mn-/mń-**. It also occurs before

some other asyllabic substantive morphemes⁶⁶: **spode łba, ze wszech miar, z(e) wszystkimi, nad(e) wsią, pod(e) drzwiami, nade wszystko but nad wszystkimi, przede wszystkim but przed wszystkimi**, etc.

Długosz-Kurczabowa & Dibisz (2001: 101) suggest that the vocalisation of prepositions has its origins in the fact that they are proclitics and historically constituted a phonetic and accentual unit with the following word, the shape of which then decided on the elision or strengthening of the preposition's yer. Today, however, the vocalisation instead depends on the phonetic shape of the following word. Vocalised prepositions occur before words beginning with consonant groups which are difficult to pronounce, and sometimes variation occurs, e.g. **w(e) trójkę, w(e) czwartek**. Forms with **we** which are of a regional character are most frequent in the south and south-west of Poland⁶⁷, e.g. **we Wiedniu, we wodzie**.

All the information given by Kowalik, as well as Długosz-Kurczabowa & Dibisz, is confirmed by the previous data, but their statements are far too general to be practically applicable.

Swan (2002b: 33f.), however, is far more detailed. He says that the "prepositions **bez, (z)nad, od, (s)pod, (s)przed, przez, w**, and **z** end in an underlying -Ø, which will be realized as mobile **e** under limited circumstances [...]: a. The preposition **w** becomes **we** before **w** or **f** followed by another consonant [...] b. The preposition **z** (in whichever meaning) becomes **ze** before **s, z, ś, ź, ż, sz, wz, ws** followed by another consonant [...] c. All of the above prepositions may take a following **e** before asyllabic surface stems⁶⁶ of nouns and pronouns, in particular including oblique forms of **ja I** beginning with **mn-**: **ze mną** *with me*, **przede mną** *in front of me*, **nade mną** *over me*, and so on. See also: **we śnie** *in the sleep*, **we łby** *in the heads*, **spode łba** *stealthily, from beneath one's brow*, **ze łzami** *with tears*, **we mchu** *in the moss*, **we mnie** *in me*, and so on[...] d. Occasionally, mobile vowels will occur in set expressions but not in regular collocations, as in **z krwi** *from one's blood* but possibly **ze krwi i kości** *in one's very bone*, as in **Polak ze krwi i kości**⁶⁸ *a Pole through and through*, **w dniu** *on the day*, but **we dniu i w nocy** *by day and night*."

The introduction of an underlying empty position is quite interesting, but the arguments for its existence are unclear. The vocalisation of **z** before ⟨wz⟩ and ⟨ws⟩ is presented as obligatory, and only occurring if these two clusters

⁶⁶ By *asyllabic substantival morphemes* or *asyllabic surface stems*, Kowalik (1999) and Swan (2002b) refer to stems which in the form under discussion do not contain any surface vowel. The only vowels in such words are found in the ending, e.g. **psa** *dog* with the stem {ps} and the accusative/genitive singular ending {a}.

⁶⁷ "[N]a obszarze południowej Małopolski i Śląska."

⁶⁸ A Google search, however, gives only 6 hits for this phrase with the vocalised form of the preposition, whereas it gives 64,400 hits with the non-vocalised form.

are in turn followed by a consonant. Previous data have shown clearly that this is not the case. Vocalisation is not obligatory in the context mentioned, but is very frequent, and this type of vocalisation frequently occurs even if the consonant clusters are followed by a vowel. The principle c) is quite vague, and does not clearly state whether the instances are optional or only represent a certain number of collocational types involving the context mentioned. The context itself also seems very spurious, as there are many instances of non-phonological vocalisation before stems containing a full vowel, e.g. **przede wszystkim** *above all* or **we Lwowie** *in Lviv*. Therefore Kowalik's (1999) reference to syllabic nominal morphemes is more appropriate, as these words can be analysed as derived from the root morphemes {vš} and {lv}. Swan's final remark about similar collocations, with and without the vocalised form of a preposition, is however quite informative.

Gruszczyński & Bralczyk (2002: 199) state that "in order to facilitate the pronunciation by avoiding consonant clusters, the primary prepositions may occur in alternative forms [...]. In southern and western Poland the prepositional forms with *-e* are used before every form beginning with a consonant, e.g. *we wodzie ze sokiem*. It is not recommended to use such forms in careful standard Polish."

This information is meagre to such a degree that it becomes completely inaccurate. The impression is given that in certain dialects any preposition may be used in the vocalised form before any word beginning with a consonant.

Bańko (2002: 114f.) lists the eleven prepositions **bez**, **nad**, **od**, **pod**, **przed**, **przez**, **spod**, **sprzed**, **znad**, **w** and **z** as having a non-vocalised and a vocalised form. He then states the following five principles: 1) The vocalised forms are regularly used before forms of the pronoun **ja** I. 2) The form **we** occurs regularly before words beginning with a consonant cluster where the first letter is ⟨w⟩ or ⟨f⟩. 3) The form **ze** is regularly used before a consonant cluster where the initial letters are ⟨z⟩, ⟨s⟩, ⟨ż⟩, ⟨ś⟩, ⟨ź⟩, ⟨rz⟩ or ⟨sz⟩. Often, but not always, **ze** also occurs before the clusters ⟨wz⟩ and ⟨ws⟩. 4) The vocalised forms occur obligatorily or optionally in certain collocations. Obligatory forms are, for instance, **nade wszystko**, **przede wszystkim**, **we wszystkim**, **we wszystko**, **ze wszech miar**, **we łbie**, **we śnie**, **we Lwowie**, **ze łzami w oczach**, **patrząc spode łba**, **dać się we znaki**, **zrobić coś we dwóch** (or **we trzech**), **fora ze dwóra**. Optional but prevailing forms are, for instance, **ze sobą**, **we trójkę**, **we krwi**, **we łzach**, **we mgle**. Optional and rare are: **we śróde**, **we czwartek**, **we mszy**⁶⁹. 5) In southern and western Poland the form **we** occurs before other words beginning with ⟨w⟩ or ⟨f⟩, and the form **ze** occurs before other words

⁶⁹ According to both the corpus data and the Google data, **we mszy** with the vocalised form of the preposition is somewhat more frequent than with the non-vocalised form.

beginning with ⟨z⟩, ⟨s⟩, ⟨ź⟩, ⟨ś⟩, ⟨ż⟩, ⟨rz⟩ or ⟨sz⟩. This usage has a regional character. Finally, Bańko adds that the approximative particle **z(e)** only vocalises before the words **dwa, dwanaście, dwadzieścia, dwieście, trzy, trzynaście, trzydzieści, trzysta, cztery, czternaście, czterdzieści, czterysta**, and **sto**. In all these contexts, except before **sto**, the non-vocalised form is also possible, but more rare.

Bańko presents a very good account of the issue. All of the prepositions involved are mentioned and the contexts for the phonologically triggered vocalisation of **w** and **z** are precisely defined. The context ⟨rz⟩ followed by a consonant, however, seems quite dubious, as this digraph represents a further development of a palatalised /r/, occurring under the influence of a following front vowel, and the dictionaries consulted contain no entries where ⟨rz⟩ is followed by a consonant. The final comment about the collocation **ze sto** *about a hundred* is interesting, as this vocalisation as well is most probably an effect of the general principles triggering phonologically conditioned vocalisations.

6.2.4. Linguistics

Within the scientific literature, few monographs on Polish phonology discuss vowel–zero alternations in prepositions at any length. The most detailed discussion is found in a few somewhat older articles. Instead, the linguistic discussion generally directs its interest towards alternations in stems, suffixes, and partially also in prefixes. The two articles that I will discuss are also exceptional in that they are based on data collected by the author by excerption of text, whereas most of the remaining literature is based rather on native speakers' intuition and ability to generalise on the basis of quite random examples.

Grappin (1944: 13ff.) gives a quite exhaustive picture of the issue, but he stresses mainly the historical aspects of the alternations, stating that a preposition is treated as a phonetic word together with its governee, and that Havlík's law⁷⁰ applies to the collocation as a whole, e.g. **pod Warszawą** (< podъ Warszawa), but **ode mnie** (< отъ мѣне). This mechanism of the mobile *-e-*, as it has traditionally been described, has however been seriously disturbed. The vocalisation of the prepositions, even if quite similar to that of the prefixes, has to be distinguished from the latter, as the prefixed compounds are more stable than the syntactic prepositional collocations, at least if they are not fixed phrases.

The prepositions have become more exposed to influences of change in the linguistic system than have the prefixes. Their form is determined by phonetic conditions, i.e. subordinated to the structure of the governee. New factors

⁷⁰ Havlík's law is discussed on p. 41.

have intervened over time and the present situation is more complicated than Havlík's law. One has to make two distinctions, firstly between the prepositions themselves, and secondly, between the forms that their governees⁷¹ may have. The prepositions that have alternate forms are eight: **bez**, **nad**, **od**, **pod**, **przed**, **przez**, **w**, **z**. The first six are monosyllabic, often stressed and not seriously altered by assimilations. Hence, these prepositions remain faithful to the old conditions, before a group of consonants as well as before a single one. One continues to say **beze mnie** *without me*, **bez ciebie** *without you*, **nade dniem** *towards daybreak*, **nad wieczorem** *towards the evening*, **ode złego** *from evil*, **od dziś** *from today*.

The last two prepositions, i.e. **w** and **z**, consist of a single consonant. Grappin says that they do not exist, so to speak, without their governee, as they are closely bound together and are subject to extensive assimilation. **W** may only become voiceless, but **z** may also leave its point of articulation. That is why these two prepositions, and particularly **z**, more than the syllabic ones, have come to defend themselves through vocalisation. This evolution has been encouraged by the stream of borrowings, not bound by any traditions, enabling them to vocalise purely on phonetic grounds: **we Francji** *in France*, **ze Szwecji** *from Sweden*. Today, the two prepositions **w** and **z** are treated in a quite different manner if they precede a group of consonants than if they precede a single consonant.

In the first case, one can suggest the following principle: before a syntactic governee beginning with more than one consonant, of which the first consonant is phonetically similar and would give rise to a total assimilation, the prepositions **w** and **z** always vocalise if the governee is a word of foreign origin, and almost always if it is an endogenous word. Compare **w szkole** *in school*, **ze szkoły** *from school*, **we Francji** *in France*, **z Francji** *from France*, **we Włoszech** *in Italy*, **z Włoch** *from Italy*, **ze strachu** *out of fear*, **w strachu** *in fear*, **ze środy** *from Wednesday*, **w środę** *on Wednesday*, **w słońcu** *in the sun*, **ze słońcem** *with the sun*, **we wrześniu** *in September*, **z września** *from September*.

If the governee begins with a single consonant the language has been unwilling to vocalise. However, with foreign words the tendency seems to be stronger: **we formie** *in the form*, **ze Sowietami** *with the Soviets* but also **ze sejmem** *with the sejm*, **ze sobą** *with oneself*.

The majority of the information given by Grappin is confirmed by the corpus data. It is however spurious whether the syllabic prepositions are often stressed and whether they actually remain faithful to the old conditions. According to the data investigated, this is only marginally true, e.g. for collocations such as **pode mną** *under me*, but this type clearly constitutes a

⁷¹ By *governee* Grappin refers to the word which is syntactically governed by the preposition, but practically speaking he means the word following the preposition.

minority in modern standard Polish. It also seems quite spurious whether foreign borrowings would be the triggering factor for the innovative type of phonologically triggered vocalisation. In the data investigated there is no evidence that vocalisation is more frequent before words of foreign origin; rather, on the contrary, there seems to be a weak tendency to use the non-vocalised form where the vocalised form would have been expected before words, and especially names, of foreign origin. The alleged examples **we formie**, **ze Sowietami** and **ze zejmem** are all quite spurious.

The article by Klemensiewiczówna (1950) mainly deals with the regional usage of the vocalised forms **we** and **ze** before governees beginning with similar consonants followed by a vowel. However, it also contains information about the standard literary usage. It is pointed out that three different phonological contexts must be recognised, namely when the governee begins with 1) a similar consonant followed by another consonant; 2) a similar consonant followed by a vowel; 3) a totally different consonant. It is made clear that in standard literary Polish, phonologically conditioned vocalisation of **w** and **z** only occurs in the first of these three categories. As a rule, the vocalised form is also used in the phrase **ze sobą** (69% vocalised versus 31% non-vocalised, in our corpus). Further, the author points out that semantically, one must recognise three different lexemes with the form **z(e)**, namely the comitative, the ablative and the approximative. **Ze** is used in its approximative meaning before the numerals **sześć** and **siedem**, as well as before compounds beginning with these numerals. Finally, lexically conditioned vocalisations of prepositions occur before a few expressions beginning with a completely different consonant from the one found in the preposition.

As to the extended regional usage of **we** and **ze**, the author gives the approximate geographical extension, namely the area along the northern coast as well as along the western and southern borders, i.e. Pomorze, Wielkopolska, Śląsk, and western and southern Małopolska. Further, she points out that in the regions where this usage occurs, the vocalised forms are only sometimes used, and the usage is, it seems, dependent on a number of different factors such as the length of the following expression, and on logical as well as rhetorical stress. There is also a statistical difference between the two prepositions, so that **ze** is more frequent than **we** in the very same contexts. She finishes by saying that she would not consider the usage of vocalised forms in the investigated context to be wrong in speech, but that they should not be used in writing.

Practically all the information in the article is very precise and completely in line with the corpus data already investigated. Only the statement that the approximative **ze** is used before **sześć** *six* and **siedem** *seven* contradicts the findings based on corpus data.

In her next article, Klemensiewicz-Bajerowa (1955) gives the historical background to vowel-zero alternations in Polish prepositions by presenting

the basic facts about the continuation of strong and weak yers. She shows that the present day alternations cannot be derived from the historical distribution of yers. The study presents rich quantitative data, both synchronic and diachronic, on vowel-zero alternations in prepositions, and four important contexts are defined: 1) **w, z** before a consonant similar to the preposition and followed by another consonant; 2) **w, z** before a consonant similar to the preposition and followed by a vowel; 3) **w, z** before any other consonant cluster; 4) any other instance, i.e. **w, z** before a single consonant not similar to the preposition, as well as all collocations with syllabic prepositions. Then three possible ways in which the realisation of the vocalised form of prepositions might have evolved are pointed out: 1) traditional, i.e. the historically expected instances; 2) analogical, i.e. use of the unmarked non-vocalised form of the prepositions instead of the historically expected vocalised form; 3) innovational, i.e. use of the vocalised form where it is not historically expected.

Klemesiewicz-Bajerowa proposes that the traditional reflexes are often supported if the vowel in the preposition is accented because the following word is monosyllabic, e.g. **spode łba, ze mną**. The analogical tendency in prepositional phrases causes the historically expected realisations of vocalised prepositions to diminish. **Nad** followed by different forms of the pronoun **wszystek** was vocalised in 95% of instances in the 15th century, in 19% in the 16th century, in 8% in the 18th century, and no vocalisations are found in the 20th century, except in the fixed phrase **nade wszystko** *more than anything*. Innovational vocalisation occurs where the realisation of a monophonemic prepositional morpheme is threatened by assimilation, which is the case in the literary language when the following word begins with a similar consonant followed by yet another consonant. When a similar consonant is followed by a vowel, this is only a threat to the preposition in some of the dialects. When preceding a similar consonant followed by another consonant, **w** and **z** are vocalised in 96% of cases in modern written Polish. When preceding a similar consonant followed by a vowel, **w** and **z** are vocalised in 9% of instances in modern written Polish, 50% in southern and western spoken standard Polish and 90% in southern and western dialects. The author also includes within the innovational type instances where the similar consonant is preceded by a sonorant, e.g. **we Lwowie, ze wschodu**, as sonorants are weak sounds that are easily lost, which would subsequently lead to assimilation of the preposition.

The article ends with a historical survey, in which old Polish, up until around 1500, lost most of its collocations with vocalised prepositions through analogy where the preposition is not stressed. In middle Polish, approx. 1500-1800, the analogy process progresses even further, but traces of new innovative forms are found. **Z** is vocalised before the pronoun **swój** in 2% of cases in the 16th century, 6% in the 17th century, and 19% in the 18th century. A large increase in vocalised forms of **w, z** takes place in modern Polish. In a

final note the author also mentions that Czech and Slovak exhibit more vocalisations than Polish, and that the innovational vocalisations can be traced to an earlier historical stage in Czech than in Polish. In a Czech text from 1789, 43% of asyllabic prepositions are already vocalised before a similar consonant followed by a vowel and 54% before a similar consonant followed by another consonant.

Again, the formulations in the article are very precise. The only dubious point is that the vocalisation in the collocation **w** **Lwowie** in *Lviv* is considered to be innovative, hence phonologically triggered. Instead, this rather seems to be an instance of traditional vocalisation, as one may argue that a historical yer is highly plausible between the two initial consonants, just as in the underlying root morpheme {lew} < *ЛѢВЬ *lion*. Also, the vocalised form is not used before derivations such as, for instance, the adjective **lwowski** *Lvovian*.

Biedrzycki (1974: 119) simply states that the prepositions **w** and **z** also have the forms **w**e and **z**e, which occur before **mnie** and **mna** as well as before many words beginning with a consonant cluster where the first consonant is /v/, /f/, /z/ or /s/.

The information in this statement is very scanty and very unclear, as the contexts in which the two prepositions vocalise are mixed together.

Laskowski (1975: 36) only makes very brief mention of the preposition **z(e)** when discussing alternations in prefixes. He states that the epenthesis of /e/ is phonologically conditioned, and that it prevents the group consisting of the preposition and a phonetically similar fricative following it from otherwise becoming simplified.

This statement, although fragmentary, is theoretically very interesting, as Laskowski states that vocalisation of prepositions is a case of epenthesis and that the triggering factor is to prevent the preposition from not being realised phonetically because the consonant cluster would otherwise be simplified.

Rubach (1977: 119) points out that in order for /e/ to be inserted after **w** and **z**, an identical consonant of the following stem must be part of a cluster. The /e/ is inserted with some variation if the consonant is only similar, rather than identical (i.e. differs in more features than [voiced], e.g. **z(e)** **Śląska** from *Silesia*, **z(e)** **Szczawna** from *Szczawno*. As e-insertion occurs, for instance, in **z** **słomy** of *straw*, this also gives support to the analysis of the phonetic [w] as underlying /ʎ/, i.e. as a back lateral consonant, in contrast with the lack of e-insertion when the similar consonant is followed by the glide [j], e.g. **w** **wiadukcie** in the *viaduct*, hence a rule for realisation of /ʎ/ as [w] must follow after the rule of e-insertion. He further points out other cases of variation, e.g. **w(e)** **mgle** in the *fog*, **w(e)** **krwi** in the *blood*, **w(e)** **dwóch** *two of us*, as well as some quite exceptional cases: **w** **mnie**, **z** **mnie**, **z** **mna**, but **w** **mnóstwie** in a *lot*, **z** **mnóstwem** with a *lot*.

This analysis is also theoretically interesting, as this author also discusses

the insertion of /e/. The only spurious piece of information is that variation is supposed to occur in the insertion of /e/ after the preposition *z* only before consonant clusters beginning with /š/, /ż/, /ś/, /ź/. There is no such evidence in the corpus data.

Rubach (1985) argues that all Polish prepositions have an underlying final yer, e.g. /zi/, which surfaces under certain circumstances and is deleted under other circumstances. He points out that there is a problem, however, because the lexical rules delete all final yers, and they must apply before the prepositional phrase has been composed by syntax. He therefore accepts the deletion and suggests another post-lexical rule of e-insertion that applies after syntax has put the prepositional phrase together. He describes the context of this rule by stating that “[e] is inserted in the preposition if the next word starts with a consonant which is identical (or similar, see below) to the consonant of the preposition. [...] the word that follows the preposition must start with a consonant cluster [...] In the dialects of southern and western Poland the requirement that there be a consonant cluster is waived. [...] the consonants in the preposition and in the next word must be similar but not necessarily identical. They may differ with respect to the value for the feature [voiced]. [...] the consonant of the preposition and the initial segment in the consonantal cluster of the following word must be fricatives and they must agree as to the values for the features [labial] and [coronal].” He also points out that before words beginning with *ś* or *ź*, “[i]f the second consonant in the word-initial cluster is a sonorant, e-insertion tends to apply in a variable manner.” He adds that the reason for e-insertion is that the yer of the prepositions *w* and *z* is deleted lexically, hence the preposition become extrasyllabic, not attached to any syllable node, and is then at risk of being assimilated.

Any other instances of vocalisation of asyllabic as well as syllabic prepositions, e.g. **we krwi**, **we dnie**, **spode łba**, are due to lexical surfacing of a final yer. Such collocations must therefore be listed in the lexicon. This is also confirmed by the fact that they normally carry idiomatic meaning. He argues that in the literal meaning we find **w krwi**, **w dni**, **spod łba**, because the semantics are entirely compositional and the preposition may be separated from the head noun by an intervening determination, hence they are generated outside the lexicon, by syntax. Sometimes other irregularities can also be observed in the lexicalisation, eg. **fora ze dwora** as opposed to **z dworu**. Some instances with regular semantics but irregular vocalisation must, however, also be listed in the lexicon: **we środę**, **ze czcią**, **we mchu**, **we mgle**, **we śnie**, but they tend to regularise in the sense that they may also appear with the non-vocalised form of the preposition. The range of variability is different with different collocations and different native speakers. **We środę**, **ze czcią** are rare, whereas **we mgle** is common. **We śnie** is optional if it means “in the dream”, but more obligatory if meaning “asleep, sleeping”. Hence forms that are weakly motivated for listing in the lexicon tend to lose their

lexical status if the sole reason for listing is phonological, and not semantic. There is however an obvious counterexample, the position before **mnie** and **ma**, where all affected prepositions vocalise. These collocations do not exhibit any tendency to lose the final /e/, nor to develop any idiomatic meaning.

Rubach's discussion of the ordering of the rules for deletion of yers and e-insertion is theoretically interesting. One wonders, however, why he postulates the prepositional final yers, if they are always deleted, and if e-insertion works without reference to the yers. The only spurious point is the reported variation in the vocalisation of **z** before /ś/ or /ź/ followed by a sonorant. There is no evidence in the corpus data that such variation of any significance exists.

Szpyra (1989: 216, 221) also states that yer deletion or strengthening applies at the word level, which is why /zī pīsamī/ gives **z psami** *with the dogs*. Vowel epenthesis then operates at phrase level, giving rise, for instance, to **ze strachu** *out of fear*.

Szpyra's reasoning means that in the latter example, the yer of the preposition is first deleted at word level in the underlying //zī straxu//, then an epenthic vowel is inserted at phrase level in order to prevent the preposition from becoming assimilated and lost in the pronunciation.

Bethin (1992: 145f.) argues that there is an interaction at play between vowel-zero alternations and syllabification. She agrees with Rubach (1985) that there is a rule of epenthesis at work in order to break up sequences of similar consonants in clusters of three or more consonant segments, which "may be motivated by the unsyllabifiability of the preposition in terms of possible syllable onsets".

Gussmann (2007: 234ff.) discusses vowel-zero alternations in prepositions at some length, and he mentions **w**, **od**, **bez**, **z**, **przez**, **pod**, **nad**, **spod** as alternating, giving examples of his own usage, e.g. **we krwi** (*running*) *in the blood*, **w krwi** *in blood*, **pode drzwiami** *by the door*, **pod drzwiami** *under the door*, **we śnie** *while asleep*, **w snach** *in dreams*. He argues that it is not possible to maintain that the vocalised form of prepositions generally occurs before heavy consonantal clusters, because of instances like **z pstrągiem** *with a trout*. It is however true that the vocalised forms of prepositions never occur before a single consonant, i.e. a two-member consonant sequence is a necessary but not sufficient condition for vocalisation of prepositions. In terms of Government Phonology (GP)⁷², he further states that the preposition itself ends with a floating melody, i.e. phonetic material that is not attached to any constituent in the syllabic structure. The following triggering word also contains a floating melody or an empty nucleus, i.e. a syllabic constituent without any phonetic material, as a floating melody becomes attached to the syllabic structure if it is

⁷² Government Phonology is discussed in more detail on p. 267ff.

followed by another unattached floating melody or an empty nucleus. A strong argument for this analysis is reportedly that very few of the onsets involved in the word following a vocalised preposition can qualify as well-formed onsets in GP, nor do they in general phonological terms.

However, in order for the vocalisation process to take place, the preposition must be able to see the structure of the following word. This is only possible if the two units constitute a single domain without an intervening boundary. This also makes sense as many expressions are set phrases or petrified collocations, which are semantically non-compositional. The prepositions may also receive stress, which also indicates that the preposition and the following word must belong to a single phonological domain. Furthermore, since semantic idiosyncrasy is a vague phenomenon, it is not surprising to find variation or vacillation. All such single domain expressions must be listed in the lexicon, but they only constitute a minor part of the collocations with vocalised prepositions.

Syntactically derived prepositional phrases must contain a domain boundary, and the majority of vocalised prepositions occur under such circumstances. Therefore, melodic considerations evidently override the domain structure, as **w** and **z** always vocalise when the following word begins with a cluster whose first consonant is similar to that of the preposition. Hence what is at stake is melodic similarity rather than segmental complexity. It is also pointed out that vocalisations triggered by the melody never occur before single consonants, nor in any of the syllabic prepositions.

Gussmann's analysis, in which the vocalising prepositions carry a final floating melody, and the presence or absence of a boundary constitutes the difference between phonologically and lexically triggered vocalisations, is theoretically highly interesting as it makes a homogeneous representation of the prepositions possible without any ordering of separate rules for deletion and insertion.

6.3. Summary of Polish Data

The findings based on the corpus data and the review of the previous literature on the subject will now be summarised in order to give an easily accessible survey of the vocalisations which occur in modern standard Polish.

Eleven prepositions exhibit two alternating forms, with or without an additional final /e/. These prepositions fall into two groups: the syllabic **bez**, **od**, **nad**, **pod**, **przed**, **przez**, **spod**, **sprzed**, **znad**, and the assyllabic **w**, **z**. It should be noted that **z** represents three lexemes: a comitative preposition governing the instrumental case, an ablative/elative/partitive preposition governing the genitive case, and an approximative adverbial particle.

The grounds for vocalisation also fall into two main groups: phonological and lexical. Only the two assyllabic prepositions vocalise on phonological

grounds, whereas vocalised forms of both types of prepositions occur on lexical grounds.

6.3.1. Phonological Vocalisations

Phonologically triggered vocalisations of the prepositions **w** and **z** occur whenever the right phonological context is present, i.e. without any other reference to the following lexeme than its phonological structure.

In modern standard Polish, phonological principles trigger the vocalisation of **w** before consonant clusters beginning with /v/ or /f/. as the same is true for the vocalisation of **z**, in whichever of its three meanings, before consonant clusters beginning with /s/, /z/, /š/, /ž/, /ś/ or /ź/. In some regions, especially in the south, a dialectally based tendency to vocalise before the above consonants can be observed even if they are not part of consonant clusters. Such usage is however considered sub-standard.

The corpus data suggest that /A/ is treated as a possible second member of a consonant cluster, whereas /j/ is not, hence in modern standard Polish we get **we własnym** /vevłasnym/ [vɛ'vwasnim] *in one's own*, but **w Wiedniu** /vɔvjednú/ [v'ɔjɛdɲu] *in Vienna*.

There is also a close to obligatory principle according to which **z** vocalises before words beginning with /vz/ or /vs/ if followed by another consonant. Further, it seems that this principle may optionally, but less frequently, be broadened to apply before any word beginning with /vs/, /vz/, /vš/, /vž/, /vś/, /vź/ irrespective of the following phoneme. It is unclear whether regional factors may possibly influence this variation.

6.3.2. Lexical Vocalisations

All other instances of vocalisation, apart from the above mentioned phonologically triggered ones, must be encoded in the lexicon of modern standard Polish.

It should especially be noted that, as a rule, in instances of lexical vocalisation the stress is transferred to the additional /e/ of the preposition if the following word is monosyllabic.

From the synchronic perspective it is not relevant whether the collocations are historically motivated or not. Therefore the attractive division into traditional, analogical and innovative, proposed by Klemensiewicz-Bajerowa (1955), is not applied here. Instead, a division based on the presence or absence of a contextual and/or semantic layer in the lexicalisation seem to be primary, and the obligatoriness of the vocalisation then serves as a secondary basis for the following classification of the collocations.

The following list of lexicalised vocalisations constitutes a summary of the information in the literature and the corpus data which have been described earlier in this study. For obvious reasons a certain amount of qualified

approximation has been necessary in order to group the collocational types according to frequency, as there are sometimes slight discrepancies between the corpus data and the statements in the literature.

6.3.2.1. Formal Lexicalised Vocalisations

Formal lexicalised vocalisations refer to instances of non-phonological vocalisation occurring in collocations where the semantics of the collocation do not differ from the sum of the semantics of its constituents and where no limitations can be observed with regard to the contexts in which the collocation may be used, i.e. the same constituents do not occur in a non-vocalising collocation with another meaning or in certain other contexts. Of course, the optional formal lexicalised vocalisations occur with the non-vocalised form as well, but these optional forms refer to the very same meaning and context.

6.3.2.1.1. Obligatory

The vocalised form is considered obligatory in the following lexicalised collocational types according to several sources:

- all the vocalising prepositions before **mną/mnie** *me*: **beze mnie, nade mną, nade mnie, ode mnie, pode mną, pode mnie, przede mną, przede mnie, przede mnie, spode mnie, sprzede mnie, we mnie, ze mną, ze mnie, znade mnie**;
- **we Lwowie** *in Lviv*, **ze Lwowa** *from Lviv*.

According to the data investigated, some of these collocations are not vocalised at a rate of 100%, which means that their obligatoriness is to a certain degree based on tradition. Actually, in the Google data the traditionally non-obligatory **we mgle** (96%) and **ze sobą** (94%) are somewhat more frequent than **we Lwowie** (94%) and **ze Lwowa** (85%), which are traditionally considered to be obligatory.

6.3.2.1.2. Prevailing

Optional, but prevailing, are:

- **ze Lwowem** *with Lviv*;
- **ze łzami** *with tears*;
- **we mgle** *in the fog*;
- **we mszy** *in the mass*;
- **ze sobą** *with oneself, with each other*;
- **ze wszech** *from all*, **ze wszystkich** *from/of all*, **ze wszystkimi** *with all*.

6.3.2.1.3. Equally Distributed

Optional and occurring as frequently as the corresponding collocation with the non-vocalised form are:

- **ze chrztu** *from the baptising*;
- **we czci** *in honour*, **ze czcią** *with honour*;
- **we mgłach** *in the fogs*;
- **ze mszą** *with the mass*.

6.3.2.1.4. Infrequent

Optional and infrequent are:

- **we czwartek** *on Thursday*, **we środe** *on Wednesday*;
- **we dworze** *in the courtyard, at the manor*;
- **ze lnu** *of flax, linen*;
- **we mchu** *in the moss*;
- **ze młyna** *from the mill*;
- **ze wsią** *with the village/countryside*;
- **ze wszelkich** *from/of all kinds of*, **ze wszelkimi** *with all kinds of*;
- **we zbożach** *in the crops*;
- **we zwyczaj** *in the habit*;
- **ode żniwa** *from the harvest*;

6.3.2.1.5. Rare

Very rare in written modern standard Polish and often reported as archaic or regional are:

- **nade dnem** *over the bottom*, **ode dna** *from the bottom*;
- **przede dniem** *before the day*;
- **nade drogą** *above the road*, **ode drogi** *from the road*, **pode drogą** *beneath the road*;
- **nade drzwiami** *above the door*, **ode drzwi** *from the door*, **przede drzwiami** *in front of the door*, **przeze drzwi** *through the door*, **sprzede drzwi** *from outside the door*;
- **ode dworu** *from the court, mansion*;
- **ode krwi** *from blood*;
- **ode Lwowa** *from Lviv*, **spode Lwowa** *from outside Lviv*;
- **nade wsią** *over the village*, **ode wsi** *from the countryside, village*.

6.3.2.2. Contextual Lexicalised Vocalisations

In this category, most vocalised collocations are clearly semantically conditioned, but not all. Some of them have a meaning that does not differ noticeably from the meaning of their constituent parts, but instead the collocation with the vocalised form is only used in certain restricted contexts. Contextual restrictions are of course always to be found in cases of semantic lexicalisations.

It is sometimes also the case that a contextual lexicalisation contains some morphological irregularities which do not exist in the corresponding

collocation with the non-vocalised form of the prepositions when used in the literal meaning of the collocation's constituent parts.

It should further be noted that contextual vocalisations with a high rate of vocalisation seem to have a tendency also to spread over to instances of the corresponding collocation used in the literal meaning of its constituent parts. It is actually not very surprising that the non-vocalised form may be perceived as odd by many native speakers, as the usage of the non-vocalised form in the literal meaning of the collocation is far less frequent, e.g. the form **we krwi** is also encountered in the literal locative meaning *in the blood*, as well as its metaphorical meaning *running in the blood*, which is synonymous with *in one's nature*.

6.3.2.2.1. Obligatory

The following collocational types may be considered obligatory:

- **we dwóch, we trzech, we czterech** (masculine), **we dwie, we trzy, we cztery** (feminine) denoting a group of people doing something together, e.g. **pracowali we trzech** *they were working three men together*, but in the literal meaning e.g. **w trzech tomach** *in three volumes*;
- **fora ze dwora** *get out of here*, as opposed to the literal meaning **z dworu** *from the courtyard*;
- **we dnie (i) w nocy** *day and night, all around the clock*;
- **we krwi** *running in the blood*;
- **(patrzeć) spode łba** *(to look) stealthily, to scowl, to glower, we łbie* *in the scull, head, mind* (of a person), as opposed to e.g. the head of a nail or a screw;
- **(Maryna) ze młyna** *(Maryna) from the mill* as a rhyming apposition to a girl's name;
- **we śnie** *sleeping, asleep*;
- **ze wsi** *from the countryside* as opposed to **z wsi** *from the village(s)*;
- **nade wszystko** *above all, przede wszystkim* *above all*;
- **zbaw nas ode złego** *deliver us from evil*;
- **dać/dawać (komuś) się we znaki** *to make life a misery (for somebody)*;
- **Pode Dworem** *(At the Manor)* as proper name of a couple of small communities.

The obligatoriness of these collocations is based either on their very high rate of vocalisation or on implicit recommendations in the literature. They are however normally not explicitly referred to as obligatory, but rather reported to have a special status as only/always being used in certain contexts.

6.3.2.2.2. Prevailing

Optional, but prevailing, are:

- the approximative **ze** before all numerals, basic as well as compound, that begin with {**dw**} 2, {**trz**} 3, or {**czter**} 4;

- **we dwa kije** (*beaten*) with two sticks;
- **we łzach** in tears, crying;
- **we łbach** in the skulls, heads, minds, **ze łba** from the skull, head, mind (of a person, as opposed to other kinds of heads, e.g. of nails);
- **we dwoje**, **we troje** (mixed gender), **we dwójkę** (unspecified gender) denoting a group of people doing something together.

6.3.2.2.3. Infrequent

Optional and infrequent are:

- **we czworo** (mixed gender), **we trójkę**, **we czwórkę** (unspecified gender) denoting a group of people doing something together
- **we dwa konie** (*hitched*) with two horses,
- the approximative **ze** before {**sze**} 6 and {**sied**} 7, possibly representing a regional usage.⁷³

6.3.2.2.4. Rare

Very rare in written modern standard Polish and partially reported as archaic or regional are:

- **we cztery oczy** *between four eyes*;
- **nade dniem** *at daybreak*;
- **pode drzwiami** *at the door(step)*, **pode drzwi** *to the door(step)*, **we drzwiach** *in the door (opening), on the doorstep*.

⁷³ Before 6 and 7, according to Klemensiewiczówna (1950), but not confirmed by other sources. However, a simple Google search on Polish pages gives the following frequencies: **ze dwa** 78%, **ze dwie** 79%, **ze trzy** 81%, **ze cztery** 72%, **ze sześć** 35%, **ze siedem** 12%.

7. The Slovak Data

The Slovak prepositions differ from their Polish and Czech counterparts in that the vowel appearing in the vocalised form is predominantly /o/. As was previously pointed out, prepositions in Slovak occur in their vocalised form to a significantly higher degree than in Polish, but still to a significantly lower degree than in Czech. The previous findings about the vocalisation of the Polish prepositions will now serve as a point of departure for the investigation of the Slovak prepositions, as some parallel patterns are to be expected; at the same time, other specifically Slovak patterns are expected in order to explain the much higher frequency of vocalisation in Slovak.

As with the Polish data, all instances of the vocalised form of the prepositions are extracted from the corpus together with the following context word, and the possibility of using the non-vocalised form in the same context is then investigated. Based on these data, some preliminary hypotheses are formulated about the usage of the different forms of the Slovak prepositions.

Next, the available literature is presented and the findings based on the corpus data are compared to the data in the literature. Similarities and differences between the corpus data and the literature are discussed, and finally, the findings are summed up.

7.1. The Corpus Data

The 15 relevant prepositions are listed in Table 37. The numbers of occurrences in the corpus of the vocalised and non-vocalised forms are given, as well as the relative frequency of the vocalised forms of the prepositions.

Preposition	Vocalised		Non-vocalised
znad(o)	0		3
spod(o)	0	0.0%	1,579
cez(o)	8	0.1%	12,216
pod(o)	34	0.1%	28,478
bez(o)	144	0.5%	29,273
spred(o)	4	0.5%	765
nad(o)	238	1.0%	23,627
od(o)	1,038	1.3%	79,490
pred(o)	746	1.5%	50,136
v(o)	70,536	8.4%	767,641
k(u)	11,849	9.8%	109,598
s(o)	45,198	15.9%	238,982
z(o)	46,719	17.0%	228,172
navzdor(y)	52	69.3%	23

skrz(e)	208	73.0%	77
Total	176,774	10.1%	1,570,060

Table 37. Vocalisation of Slovak prepositions.

The prepositions have been ordered according to their relative rate of vocalisation in the corpus. They clearly fall into three different groups: syllabic prepositions with very low frequencies of the vocalised form (below 2%), asyllabic prepositions with a relative frequency between 8% and 17%, and syllabic prepositions with a frequency of the vocalised form around 70%. The absolute frequency of this last type of preposition in either form is however low, and they also have in common that no similar prepositions occur in the Polish data, which points to the fact that they do not constitute part of the core of Slovak prepositions. Also, they do not vocalise by means of the vowel /o/, which is otherwise used in the majority of Slovak prepositions.

Just as with the Polish data, the investigation will begin with the syllabic prepositions and then continue to the asyllabic ones.

7.1.1. Syllabic Prepositions

As a general observation, the syllabic prepositions are found in collocations with only a restricted number of word forms each.

7.1.1.1. bez(o)

Already when it comes to this first preposition, there is a noticeable difference between the Slovak and the Polish data. In Polish, the preposition **bez** vocalises only in one single collocation. In Slovak, a somewhat wider range of different collocations occur in the corpus. With spelling mistakes such as **bezo mňy** for **bezo mňa** *without me* omitted, the following collocations with the vocalised form of the preposition **bez(o)** were found in the corpus.

bezo dna	2	bez dna	16
bezo krýdel	1	bez krídel	5
		bez krídiel	4
bezo mňa	76	bez mňa	5
bezo slov	4	bez slov	136
bezo slova	1	bez slova	581
bezo sna	3	bez sna	10
bezo sporu	7	bez sporu	1
bezo starosti	1	bez starosti	55
		bez starostí	9
bezo strát	1	bez strát	16
bezo špásov	1	bez špásu	1
bezo špaty	1	bez špaty	0

		bez šp*	31
bezo tužky	1	bez tužky	0
		bez tu*	22
bezo vln	1	bez vln	0
		bez vlniek	2
		bez vlny	5
bezo všetkej	1	bez všetkej	86
		bez všetk*	296
bezo zbrane	1	bez zbrane	12
		bez zbrani	5
		bez zbraní	2
bezo zbytku	2	bez zbytku	0
		bez zbytočn*	86
bezo zmeny	13	bez zmeny	42
bezo zmien	2	bez zmien	31
bezo zvláštnej	1	bez zvláštnej	9
		bez zvláštn*	45
bezo zvyšku	22	bez zvyšku	13

Table 38. Vocalisation of the preposition **bez** in the Slovak corpus.

All single instances as well as the two instances of **bezo dna** *without bottom* and the three instances of **bezo sna** *without sleep* were found in fiction from the late 19th and early 20th century. Ten of the instances were found in poetry (Jesen-ský 4x, Milkin 1x, Poľský 5x) and five in prose (Kukučín 3x, Royová 1x, Rázus 1x). The single instance of **bezo zbrane** *without weapon* occurs in the contemporary press, but in a fragment which constitutes a citation from 1927. Finally, all four instances of **bezo slov** *without words* occur in the contemporary press, however they are all in citations of the title of a chapter in a newly published book with religious content. Hence all single instances are occurrences of presumably individual and/or old fashioned usage and they will therefore not be subject to further investigation. The only interesting observation so far is that most of these old fashioned instances occur before words beginning with a sibilant initial consonant cluster. A phonologically natural regularity may be traced here, as the preposition itself ends in a sibilant, but it cannot be considered part of contemporary Slovak, as it originates from a time when Slovak was still developing and when there was still a considerable influence from Czech. It becomes evident from Table 38, however, that this type of freely occurring vocalisation of **bez(o)** has been quite frequent.

The remaining collocations have a somewhat higher frequency and occur in contemporary texts. **Bezo mňa** *without me* has the highest number of occurrences (76/5). All five instances with the non-vocalised form of the

preposition occur in the old texts and three of them in poetry. In the contemporary language, **bezo mňa** seems to be an instance of obligatory vocalisation before the 1st person singular personal pronoun, just as in Polish.

There are also two other collocational types in which the vocalised form seems to be obligatory. All occurrences with the vocalised form in the collocation **bezo sporu** (7/1) *without dispute* originate from the contemporary press. The single contradictory occurrence is found in Bujnák (1919). Interestingly enough, a further 23 instances of the collocation were found written as one word, i.e. **bezosporu**, which provides further evidence for the lexicalisation of this collocation. The collocation **bezo zbytku** (2/0) *completely* (lit. *without remainder*), occurring twice in modern press, might also be obligatory. For the synonymous **bezo zvýšku** (22/13) *completely* (lit. *without remainder*), five out of the 13 instances of the non-vocalised collocational type were found in the century-old texts, whereas all 22 occurrences with the vocalised form were found in contemporary texts, which shows that preferences are clearly in favour of the collocation with the vocalised form. Besides that, a further 4 instances of **bezozvýšku** as one word also occur in the corpus, which provides further evidence for the lexicalisation of this collocation.

Both **bezo zmeny** (13/42) *without any change*, occurring six times in Jesenský's century-old texts and seven times in the contemporary press, as well as **bezo zmien** (2/31) *without changes*, occurring two times in contemporary press, should most probably be regarded as less frequent, optional vocalisations. No semantic differentiation can be traced.

It seems highly unlikely that the vocalisation of this preposition can be triggered by any phonological principles, since the non-vocalised form of the preposition prevails before practically all phonological onsets that occur in the collocations investigated. Hence the instances of vocalisation of the preposition **bez** are best explained as lexicalised historical reminiscences from a stage of the language when such vocalisations were still triggered phonologically, and as we shall see further on, they can at the same time be interpreted as having been influenced by similar collocations which occur in Czech.

7.1.1.2. cez(o)

cezo mlyn	1	cez mlyn	3
cezo mňa	7	cez mňa	1

Table 39. Vocalisation of the preposition **cez** in the Slovak corpus.

In the case of **cez(o)**, the vocalised form of the preposition is only found in two collocation types. The collocation **cezo mlyn** (1/3) *by the mill* is found in the

story *Tichá voda* by Kukučín (1892), whereas the non-vocalised form of the preposition is the only form occurring in modern texts. In the case of **cezo mňa** (7/1) *through me*, it comes as no surprise that the vocalised form dominates before the 1st person personal pronoun. The non-vocalised form **cez mňa** is found in the story *To, čo večne trvá* by Royová.⁷⁴

7.1.1.3. nad(o)

nado dňom	3	nad dňom	0
		nad dňami	1
		nad dnešn*	6
nado dverami	1	nad dverami	11
nado dvere	5	nad dvere	3
nado dvermi	24	nad dvermi	7
nado mieru	1	nad mieru	12
nado mňa	8	nad mňa	1
nado mnou	176	nad mnou	8
nado všetko	11	nad všetko	26
nado všetky	6	nad všetky	64
nado všetkých	3	nad všetkých	20

Table 40. Vocalisation of the preposition **nad** in the Slovak corpus.

The most frequent context of the vocalised form of this preposition is again before the 1st person singular personal pronoun, where it is evidently obligatory in the contemporary language. The single instance of **nad mňa** occurs in poetry in the older part of the corpus in the somewhat exceptional and rather longer context of **z nad mňa – z popod mňa** *from above me – from below me*, where two prepositions are contrasted. All eight occurrences of **nad mnou** *above me* are also found in the older part of the corpus. Two are found in metatext discussing the form of this preposition, five instances in poetry by Jesenský (1874–1945), and one instance in poetry by Tajovský (1874–1940).

The collocation **nado všetko** (11/26) is lexicalised and synonymous with **predovšetkým**. According to the present codification of Slovak, **nadovšetko** should be written as one word, and only one of the instances occurs in the contemporary part of the corpus. Written as one word, **nadovšetko** occurs 474 times in the corpus, of which 123 occurrences are found in the contemporary press, 321 occurrences in older fiction and 30 in contemporary factual texts. As

⁷⁴ First published in Czech translation in 1901. The Slovak original was lost. Retranslated into Slovak and published in 1970. It might therefore be an instance of the translator's wish to create an old-fashioned impression.

the older texts constitute a smaller part of the corpus, it becomes evident that this word is less frequent today, most probably with **predovšetkým** being used instead.

In the modern part of the corpus, the collocation **nad všetko**, with the non-vocalised form of the preposition, is primarily used when some reference, although abstract, is made to a position above something else, as in **Nad všetko však vyčnieva ich triumf...** *Above everything their triumph stands out...* or **Hráči povýšili futbal nad všetko** *The players have promoted football [to a position] above everything*. It also occurs when the pronoun **všetko** functions as an attribute to a following noun or adjective, e.g. **nad všetko očakávanie podarená** *successful above all expectation*.

The historically motivated vocalisations in **nado všetky** (6/64) and **nado všetkých** (3/20) *above all* (pl.) are all but one found in the older part of the corpus. The only modern occurrence is found in a religious text where **nado** occurs in a citation from an older text.

The three instances of **nado dňom** *at daybreak* are found in a text by Kukučín (1860-1928). All the instances of **nado dverami**, **nado dvere**, **nado dvermi** *above the door* are found in the older part of the corpus, and out of the 30 instances, as many as 20 are found in texts by Kukučín.

The single instance of the collocation **nado mieru** *extremely* (lit. *above measure*), constituting an adverbial of degree, is found in poetry by Milkin (1864-1920).

7.1.1.4. od(o)

In the corpus data, a few typing mistakes were detected, such as **odo mnňa**, **odo nňa** for **odo mňa**, as well as an occurrence of a dialectal **odo mňe**. Such occurrences are omitted from the data in the following table.

odo dna	10	od dna	27
odo dňa	121	od dňa	23
odo dnes	8	od dnes	79
odo dneška	23	od dneška	22
odo dverí	64	od dverí	31
odo dvier	11	od dvier	34
odo dvora	14	od dvora	16
odo dvoríka	1	od dvoríka	0
odo mňa	781	od mňa	8
odo stáda	1	od stáda	0
		od st*	1511
odo všetkých	1	od všetkých	168
		od vš*	290

Table 41. Vocalisation of the preposition **od** in the Slovak corpus.

Once again, the majority of instances of the vocalised form of the preposition occur with the 1st person singular personal pronoun in the collocational type **odo mňa** (781/8) *from me*. This usage seems obligatory in contemporary standard Slovak, and seven of the eight contrasting instances of **od mňa** occur in the older part of the corpus. One occurrence is found in a contemporary, religious text.

The second most frequent collocational type is **odo dňa** (121/23) *from the day*. No patterns in the usage of the two forms of the preposition could be identified with the form **dňa**. It therefore seems that both forms are employed, but that the vocalised form is far more frequent. When it comes to **odo dnes** (8/79) and **odo dneška** (23/22), both meaning *from today*, the non-vocalised form of the preposition is the most frequent in contemporary Slovak, because as many as 18 instances of **odo dneška** occur in older texts. It can also be noted that these two collocational types, with the vocalised form of the preposition, are never followed by any further complement of the type **od dnes do soboty** *from today until Saturday*, whereas the majority of instances of **od dnes** occur with some kind of further such complement.

Concerning the collocational types **odo dverí** (64/31) or **odo dvier** (11/34) *from the door*, no instance of **odo dvier** and only one single instance of **odo dverí** is encountered in contemporary texts. All the occurrences of **od dvier** are found in the older texts, and it therefore seems that the only collocational type in active use in the contemporary language is **od dverí**.

The collocational type **odo dna** (10/27) occurs 8 times in older texts, and twice in the contemporary press, both times in the context **odrážať sa odo dna** *to hit the bottom, to take off from the bottom*. On the other hand, there are 15 instances of the contrasting **odraziť/odrážať sa od dna**.

Finally, all the instances of the collocations **odo dvora** (14/16), **odo dvoríka** (1/0) *from the courtyard*, **odo stáda** (1/0) *from the herd* and **odo všetkých** (1/168) *from all* with the vocalised form of the preposition occur in the older texts.

7.1.1.5. pod(o)

podo mňa	3	pod mňa	0
podo mnou	31	pod mnou	2
		pod mn*	8

Table 42. Vocalisation of the preposition **pod** in the Slovak corpus.

The two occurrences of **pod mnou** *under me* are found in poetry in the older parts of the corpus.

7.1.1.6. pred(o)

One occurrence of a dialectal **predo všetkým** *above all*, as well as the compound **predo-zadná** *the second last*, was found in the corpus. These occurrences are omitted from the data in the following table.

predo dňom	5	pred dňom	29
predo dvanástou	1	pred dvanástou	17
predo dverami	2	pred dverami	83
predo dvercami	4	pred dvercami	0
predo dverce	1	pred dverce	0
predo dvere	79	pred dvere	24
predo dvermi	214	pred dvermi	11
predo mňa	27	pred mňa	0
predo mnou	409	pred mnou	7
predo ňou	1	pred ňou	626
predo tchormi	1	pred tchormi	0
		pred tch*	0

Table 43. Vocalisation of the preposition **pred** in the Slovak corpus.

Once again, the most frequent collocational type with the vocalised preposition is with the 1st person singular personal pronoun. Of the seven instances of **pred mnou** *before me*, six instances were found in the older texts and one instance in a contemporary translation from French of an older religious text, hence the usage of the vocalised form seems obligatory in contemporary Slovak.

The second most frequent context is in collocations with different forms of the lemma **dvere** *door*; however, all the instances of **predo dvere** (79/24) are found in the older texts, of which 64 instances come from one single author, Kukučín (1860–1928). Furthermore, all the instances of **pred dvermi** (11/214) with the non-vocalised form of the preposition are found in the older texts. All the instances of **predo dvercami** (4/0) are also found in texts by Kukučín, and **predo dverami** (2/83), **predo dverce** (1/0) in a text by Rázus (1888–1937). This means that the only three collocational types occurring in the contemporary texts are **predo dvermi**, **pred dverami** and **pred dvere**. The forms used with the non-vocalised forms of the prepositions are the regular forms in the instrumental and accusative cases, whereas the instrumental form **dvermi** is irregular. Out of the 214 instances of **predo dvermi**, only 45 occur in the contemporary texts, and all but one of these instances in modern Slovak have a temporal meaning, as in e.g. **voľby sú už pred dvermi** *elections are around the corner* (lit. *elections are already at the door*). Only once is this collocation used in its spatial meaning: **predo dvermi stál Karel** *Karel stood outside the door*. The contrasting collocational type **pred dverami**, with the non-vocalised form of

the preposition, is equally used in temporal and spatial meaning in the modern language, whereas in the older texts **predo dvermi** is also used both with temporal and spatial meaning.

All instances of the collocational types **predo dňom** (5/29) *before the day*, **predo dvanástou** (1/17) *before twelve o'clock*, **predo ňou** (1/626) *in front of her*, and **predo tchormi** (1/0) *from the polecats*⁷⁵ occur in the older part of the corpus, and furthermore, the last two collocational types are found in poetry.

7.1.1.7. spred(o)

spredo dverí	2	spred dverí	5
spredo dveriec	1	spred dveriec	0
spredo dvier	1	spred dvier	0

Table 44. Vocalisation of the preposition **spred** in the Slovak corpus.

All the occurrences of the vocalised form of this preposition occur in the older part of the corpus. The expected collocational type **spredo mňa** *from in front of me* does not occur in the corpus, nor does the contrasting **spred mňa**. Searches were also made for the older spelling **zpred(o)** without result.

7.1.1.8. spod(o), znad(o)

Although the form **znado** *from above* is listed in most dictionaries and **spodo** *from under* is easily retrieved on the Internet, these two vocalised forms were not encountered at all in the corpus data. Searches were also made for the older spelling **zpodo**, but without result.

7.1.1.9. navzdor(y)

The preposition **navzdor(y)** *despite* is not very frequent in Slovak and its structure and behaviour differ from the other syllabic prepositions in many ways. First of all it is worth mentioning that the root of this preposition is the noun **vzdor** *defiance*, and the final alternating vowel is probably the plural ending *-y*, which may be added to this noun. It is used not only as a preposition, but also as a postposition, e.g. **víchrom navzdory** *despite the strong winds*. Such postpositional instances, as well as one occurrence in Czech, have been omitted in the following table.

navzdory b*	1		
navzdory č*	1	navzdor č*	1
navzdory d*	1		

⁷⁵ The translation is due to occurrence of this collocation after the verb **hájim** *I protect*.

navzdory e*	1		
navzdory f*	1		
navzdory h*	1		
navzdory j*	1	navzdor j*	2
navzdory l*	1		
navzdory l̃*	1		
navzdory m*	2		
navzdory n*	4	navzdor n*	2
navzdory o*	2	navzdor o*	1
navzdory p*	4		
navzdory r*	2		
navzdory s*	3	navzdor s*	2
navzdory t*	8	navzdor t*	5
navzdory t̃*	1		
navzdory ú*	2		
navzdory v*	8	navzdor v*	4
navzdory z*	1	navzdor z*	1

Table 45. Vocalisation of the preposition **navzdor** in the Slovak corpus.

No systematic pattern could be traced in the usage of the two forms of this preposition. The vowel–zero alternation in this preposition must therefore be considered an instance of free variation with a clearly noticeable tendency towards a preference for the longer, vocalised form.

7.1.1.10. **skrz(e)**

Just as with **navzdor(y)**, there is not much to be said about the preposition **skrz(e)** *through*. It is not very frequent, is basically synonymous with the preposition **cez(o)**, and may be considered a borrowing from Czech. One overt sign of its oddity is that the alternating vowel is not **-o** as in the majority of Slovak prepositions, but **-e**, as in the majority of Czech and Polish prepositions.

The corpus data contain quite a few instances of this word form that are not instances of the Slovak preposition **skrz(e)**, such as adverbial usage of the same lexeme without any following noun phrase, e.g. **idú skrz** *they go through*, one instance of **skrz** as a postposition, i.e. **dosku skrz** *through the plank*, the abbreviation **SKrZ**, the web address **www.skrz.sk**, and as many as 30 instances of **skrze** and seven instances of **skrz** that actually occur in Czech sentences, which gives an indication that this preposition is more frequent in Czech than in Slovak. All such occurrences have been omitted from the following data.

skrze a*	3		
skrze b*	4	skrz b*	1
skrze c*	3	skrz c*	1
skrze d*	9	skrz d*	1
skrze f*	2	skrz f*	1
skrze h*	3	skrz h*	1
skrze j*	9		
skrze k*	13	skrz k*	4
skrze l*	4	skrz l*	1
skrze l̄*	1	skrz l̄*	1
skrze mňa	4	skrz mňa	0
skrze m*	8	skrz m*	1
skrze n*	21	skrz n*	5
skrze ň*	3	skrz ň*	1
skrze o*	7	skrz o*	1
skrze p*	9	skrz p*	2
skrze r*	4	skrz r*	1
skrze s*	32	skrz s*	2
skrze t*	23	skrz t*	9
skrze u*	3	skrz u*	2
skrze v*	11	skrz v*	5
skrze z*	2	skrz z*	2
skrze ž*	2	skrz ž*	1

Table 46. Vocalisation of the preposition **skrz** in the Slovak corpus.

Once again, no systematic pattern can be traced in the usage of the two forms. The vowel–zero alternation in this preposition must therefore be considered an instance of free variation with a very clear tendency towards a preference for the longer, vocalised form, as actually only three of the instances of the non-vocalised form occur in contemporary texts, and one of them is a citation from the Bible. Furthermore, only 34 instances of **skrze** are found in contemporary texts, and as many as 25 of those instances occur in a religious context, e.g. **spásy skrze Ježiša Krista** *salvation through Jesus Christ*. Furthermore, within the press texts it occurs in only one of the newspapers, *Slovenská republika*, with as many as 21 instances, but not in any of the other newspapers and magazines. Finally, it should be noticed that this preposition is never used in the modern part of the corpus in a spatial meaning. In most instances it is used to express different kinds of means, hence this preposition seems somewhat archaic, and in modern Slovak there is a clear tendency towards the vocalised form.

7.1.2. Asyllabic Prepositions

Just as in the case of Polish, the working hypothesis is that the Slovak asyllabic prepositions are likely to occur in their vocalised form before phonological strings, together with which the purely consonantal prepositions would otherwise give rise to phonological structures that are disallowed by the phonological principles of Slovak.

During the investigation of the asyllabic prepositions in Slovak, it turns out that the Slovak corpus contains many problematic instances of **k**, **s**, **v**, **z**, as they are not occurrences of the prepositions, but constitute initials, parts of abbreviations, or parts of headlines printed with spaces between each letter. In order to exclude this type of irrelevant data from the concordance listings, the search strings are narrowed so that the programme retrieves only instances of **k**, **s**, **v**, **z** followed either by a space and at least two letters of the alphabet or by a space and a digit. Furthermore, all instances of **k**, **s**, **v**, **z** directly followed by a full stop are excluded from the search results.

7.1.2.1. k(u)

In spite of the spelling, this preposition is phonologically /g/ or /gu/, e.g. **ku garáži** /gu garáži/ [gu'gara:ʒi] *towards the garage*. The only exception is before the personal pronouns, where it is /k/ or /ku/, e.g. **k nám** /k nám/ ['kna:m] *to us* (Pauliny 1971: 22). Of course, normal assimilation processes will in many cases generate a different pronunciation which corresponds to the spelling, e.g. **k tomu** /g tomu/ ['ktomu] *to this*.

All instances of the vocalised preposition **ku** were extracted and sorted with respect to the first grapheme of the following word. Then all occurrences of the non-vocalised preposition **k** in the same contexts were counted for comparison. During this process a few instances were encountered that did not represent instances of these prepositions or of a normal usage as prepositions. Such instances include occurrences of the form **ku** in meta-discussion of this very form, occurrences of the abbreviation **KU** (mainly for **Katolícka univerzita**), as well as line breaks of the type **chodní- ku pavement**. Such misleading occurrences have been omitted from the data in the following table.

ku a*	21	2%	k a*	1,227
ku á*	1		k á*	17
ku b*	118	5%	k b*	2,304
ku c*	208	15%	k c*	1,151
ku č*	36	3%	k č*	1,216
ku d*	454	6%	k d*	6,557
ku d**	6	1%	k d**	461
ku e*	10	1%	k e*	785

ku f*	27	2%	k f*	1,150
ku g*	337	97%	k g*	10
ku h*	156	8%	k h*	1,906
ku i*	14	1%	k i*	1,194
ku j*	65	3%	k j*	2,461
ku k*	5,648	100%	k k*	13
ku l*	29	2%	k l*	1,248
ku l̥*	3	1%	k l̥*	409
ku m*	1,421	25%	k m*	4,234
ku n*	251	2%	k n*	16,277
ku ŋ*	3		k ŋ*	5
ku o*	88	2%	k o*	4,230
ku ô*	1		k ô*	10
ku p*	488	5%	k p*	9,546
ku r*	78	2%	k r*	3,434
ku s*	1,150	10%	k s*	10,543
ku š*	187	20%	k š*	854
ku t*	185	2%	k t*	11,251
ku t̥*	2		k t̥*	34
ku u*	9	1%	k u*	1,325
ku ú*	2	0%	k ú*	1,258
ku v*	451	6%	k v*	7,063
ku w*	2		k w*	26
ku z*	214	4%	k z*	4,922
ku ž*	29	3%	k ž*	1,011
ku 1*	12	2%	k 1*	665
ku 2*	10	5%	k 2*	206
ku 3*	14	2%	k 3*	578
ku 4*	12		k 4*	41
ku 5*	5	4%	k 5*	119
ku 6*	1		k 6*	34
ku 9*	1		k 9*	27

Table 47. Vocalisation of the preposition **k** in the Slovak corpus.

In contrast to the data concerning the vocalised forms of the Polish asyllabic prepositions **w(e)** and **z(e)**, the data concerning the Slovak **k(u)** is not as straightforward. There are small amounts of vocalisations before words beginning with almost any letter, even before words beginning with a vowel. This is quite unexpected, as no other Polish or Slovak preposition investigated so far has exhibited vocalisation before vowels.

However, quite in accordance with the findings concerning the Polish

asyllabic prepositions, the greatest numbers of the vocalised form **ku** are found before words beginning with a voiceless ⟨k⟩ and its voiced counterpart ⟨g⟩. In the same context, however, a few occurrences of **k** were also found, but out of the ten instances before ⟨g⟩, two occurrences are actually initials, five instances are occurrences before foreign names where ⟨g⟩ is pronounced [dʒ], and finally, one instance is actually a citation in old Czech where ⟨g⟩ represents /j/. This leaves only two genuine occurrences which contradict the general tendency, namely **k globálnemu boju** *to the global fight* and **k goebbelsovským metódam** *to Goebbels' methods*, both in the contemporary press. Before ⟨k⟩, nine of the 13 instances with **k** are initials, and one occurrence is a citation in Czech. The remaining three contrasting instances occur in **prišiel k kostrbu** *he came to the dead tree*, **z lásky k kvetom** *out of love towards flowers*, both in older fiction, and **k koryte rieky** *to the riverbed*, in the contemporary press.

We can therefore conclude that the vocalisation of **k** to **ku** before words beginning with /k/ and /g/ occurs in practically 100% of the instances, and in this context the contrasting contemporary occurrences of the non-vocalised form of the preposition can be suspected to be simply typing mistakes.

It has already been mentioned that the vocalised form of this preposition occurs in almost any context, even before words beginning with a vowel. The situation seems rather chaotic. Interestingly, however, it is already quite evident after a brief inspection of the data, that the majority of vocalisations in contexts other than before /k/ or /g/ occur in the older texts from the late 19th and early 20th century. These older texts constitute 28% of the Slovak corpus, but 76% of the occurrences with **ku** in contexts other than before /k/ or /g/ are found in these texts. Therefore, the older texts are disregarded in the following investigation of the contemporary distribution of **ku**. As the context words are far too many to be thoroughly investigated, only collocations with at least three occurrences in the contemporary part of the corpus have been retrieved and listed below. Instances of **ku** occurring before ⟨c⟩, pronounced as /k/, such as **ku Clintonovi** *to Clinton*, are also omitted from the following table.

ku cti	52	k cti	1
ku dnešnému dňu	6	k dnešnému dňu	53
ku dňom	3	k dňom	2
ku dňu	66	k dňu	0
ku dnu	32	k dnu	0
ku dverám	11	k dverám	21
ku hrám	3	k hrám	4
ku hre	3	k hre	25
ku hranici	4	k hranici	25
ku hviezdám	12	k hviezdám	10

ku chate	4	k chate	5
ku jednej	8	k jednej	55
ku mne	225	k mne	0
ku nám	8	k nám	1,036
ku skladbe	4	k skladbe	5
ku skončeníu	4	k skončeníu	11
ku skráteniu	3	k skráteniu	6
ku skupine	8	k skupine	15
ku skupinke	3	k skupinke	0
ku skutočnej	6	k skutočnej	5
ku skutočnému	7	k skutočnému	14
ku skutočnosti	7	k skutočnosti	7
ku skvalitneniu	6	k skvalitneniu	18
ku Slovensku	8	k Slovensku	132
ku snímkam	15	k snímkam	59
ku snímke	11	k snímke	98
ku sv.	8	k sv.	4
ku svätej	3	k svätej	5
ku svätému	3	k svätému	9
ku svätosti	5	k svätosti	24
ku svojej	30	k svojej	142
ku svojim	25	k svojim	197
ku svojmu	20	k svojmu	262
ku škodám	6	k škodám	3
ku škode	16	k škode	1
ku školám	3	k školám	1
ku špičke	3	k špičke	3
ku šťastiu	6	k šťastiu	10
ku štátu	4	k štátu	31
ku štyrom	6	k štyrom	14
ku trom	9	k trom	55
ku vám	4	k vám	113
ku včerajšiemu dňu	9	k včerajšiemu dňu	31
ku vchodu	4	k vchodu	1
ku všetkému	30	k všetkému	11
ku všetkým	66	k všetkým	32
ku vzdelaniu	4	k vzdelaniu	9
ku vzkrieseniu	3	k vzkrieseniu	0
ku vzniku	9	k vzniku	84
ku vzťahu	3	k vzťahu	6
ku zmenám	3	k zmenám	99

ku zmene	8	k zmene	237
ku zneužitiu	4	k zneužitiu	17
ku zrážke	7	k zrážke	25
ku zvieratám	4	k zvieratám	15

Table 48. Vocalisation of the preposition **k** not before /k/ or /g/ occurring at least three times in modern part of the Slovak corpus.

Only **ku** occurs in the collocations **ku dnu** *toward the bottom*, **ku dňu** *to the day*, **ku mne**⁷⁶ *to me*, and it is clearly the prevailing form in the collocations **ku cti** *to the honour* and **ku škode** *to the disadvantage*. The same is true for **ku vchodu** *towards the entrance*, **ku vzkrieseniu** *to the resurrection*, **ku školám** *towards the schools*, **ku skupinke** *towards the group*, but due to low frequencies, it is much more uncertain if these proportions are significant. Furthermore, **ku** is somewhat more frequent than **k** in collocations with **všetkému** *all*, **všetkým** *all, everybody*.

In collocations with **hviezdam** *stars*, both forms of the preposition **k(u)** are almost equally frequent. It seems, however, that **ku hviezdám** *towards the stars* is preferred in the more spatial adverbial sense, referring to an unreachable distance, whereas **k hviezdám** *to the stars* is preferred when referring to famous people.

Quite frequent is also **ku dverám** (11/21) *towards the door*, but as with the rest of the collocations in Table 48 involving nouns, it is not possible to detect any significant pattern in the choice of form of the preposition.

A context which is represented to quite a high degree in the above table, and which could reasonably be expected to trigger the vocalisation of **k**, are words beginning with a consonant cluster containing ⟨k⟩, but not beginning with it. Such initial clusters and the frequency of the two forms of the preposition are listed below.

ku sk*	96	k sk*	144
ku šk*	39	k šk*	25
ku zk*	0	k zk*	0
ku vk*	4	k vk*	1
ku vzk*	4	k vzk*	0
Total	143	Total	170

Table 49. Vocalisation of the preposition **k** before consonant clusters with ⟨k⟩ not in the first position.

⁷⁶ The prescribed, exceptional pronunciation [ˈkumn̩ɛ] is also worth noting as yet another sign of the exceptionality of this collocation.

As can be seen, they all begin with a fricative consonant. It can be noted that the relative frequency in this context is 46%, whereas the contemporary frequency of the vocalised form in contexts other than before ⟨k⟩ and ⟨g⟩ is 2%.

Quite on the contrary though, there seems to be a very interesting difference in the usage of the forms of the preposition before numerals. There are a few occurrences of **ku jednej** *to one*, **ku trom** *to three*, and **ku štyrom** *to four* in the above table, but their frequency is not very high. A closer look at these instances, as well as the instances of **ku** followed by digits, show an evident tendency towards the usage of **ku** when expressing a relationship between two numbers, e.g. **6 ku 6 hlasov** *6 against 6 votes*, **63 ku 57 hlasom** *63 against 57 votes* or **sú šance 1 ku 49 miliónom** *the chances are 1 to 49 million*. This is the case in 38 of the 55 collocations of **ku** followed by a digit. It should also be noted that **ku** in this usage is not necessarily followed by the dative case, as is otherwise always the case with this preposition. In **6 ku 6 hlasov** (*šesť ku šesť hlasov*), it rather has to be interpreted as being followed by the accusative. A search for instances of the non-vocalised form **k** both preceded and followed by digits gave only seven instances out of the 1700 occurrences with **k** followed by a digit. It therefore seems that **ku** is heavily preferred when expressing a relationship between two different numbers.

Apart from the collocations in Table 48 above, there are about 450 other instances of collocations with words which do not begin with /k/ or /g/, each occurring only once or twice, such as **ku jeho prosperite** *to his prosperity*, **ku postoju cirkvi** *towards the position of the church*, **ku sopke** *towards the volcano*, or **ku úpravám** *to the adjustments*. These are all just sporadic instances, either before a vowel, before a single consonant or before a consonant cluster. The only reasonable conclusion to be derived from these facts is that there is some sort of a tendency towards free variation between **k** and **ku** in all contexts that do not obligatorily require the vocalised form **ku**. This could also be compared to the situation in Polish, where the vocalised form **ku** has become lexicalised as the only existing form of this preposition in the modern language.

7.1.2.2. s(o)

Initially, it ought to be pointed out that this preposition, similarly to **k(u)**, has a pronunciation that differs from the orthography, and it should phonologically be represented as /z/, /zo/ in all instances except before the personal pronouns, i.e. **so syrom** /zo sirom/ [ˈzɔsɪrɔm] *with cheese*, but **so mnou** /so mnou/ [ˈsɔmnɔu] *with me*. This means that there is no phonological difference between the prepositions **s(o)** and **z(o)**, except before the personal pronouns (Pauliny 1971: 22).

First, all instances of the form **so** were extracted from the corpus. Many occurrences proved to be instances of other lexemes than the Slovak preposition **s(o)**. There were five instances of the abbreviation **SO**, 25 instances of the English lexeme **so**, 27 instances of the German lexeme **so**, several

instances of printing mistakes where **so** occurs instead of the Slovak lexemes **sa** (11x), **si** (1x), **do** (2x), **som** (4x). A few instances of typing mistakes in the word following **so** also occurred, e.g. **so ilinským** instead of the correct **so žilinským**. Furthermore, a small number of vocalisations occurred before foreign names and expressions in which the first letter diverges from the principles of Slovak pronunciation, e.g. **so Cindy**, **so Cher**, **so Jacquesom**, **so chargé d'affaires**.

Secondly, contrasting data with the non-vocalised form of the preposition in the same contexts were retrieved from the corpus. Some misleading data consisting of occurrences not representing the preposition **s** were also identified here, such as a few printing mistakes where **s** occurred instead of **sa** (1x), **v** (5x), **a** (7x). There were also more than 100 instances of a single **s** or **S** being an initial, an abbreviation for **sekunda** or an English genitive suffix, most frequently in the collocation **McDonald's**. Finally, the non-vocalised form of the preposition occurred almost 250 times before different abbreviations beginning with an **S** or **Š**, which as a rule are spelled out so that the initial letter is pronounced [ɛs]/[ɛz] and [ɛj]/[ɛʒ] respectively, hence e.g. **s SBA** is pronounced as [zɛzβɛ'a:]. Finally there were a handful of completely uninterpretable occurrences and a few occurrences in Czech.

All such misleading data have been omitted in the following investigation.

so b*	1	0%	s b*	6,339
so c*	40	1%	s c*	5,599
so č*	1	0%	s č*	3,804
so d*	112	1%	s d*	10,692
so h*	1	0%	s h*	5,483
so l*	6	0%	s l*	3,379
so m*	1,969	12%	s m*	13,994
so o*	1	0%	s o*	11,450
so p*	201	1%	s p*	29,637
so s*	22,846	100%	s s*	36
so š*	3,796	99%	s š*	21
so t*	20	0%	s t*	20,167
so v*	2,161	11%	s v*	17,498
so z*	10,403	100%	s z*	9
so ž*	2,225	99%	s ž*	30
so 1*	314	35%	s 1*	596
so 2*	5	1%	s 2*	585
so 3*	3	1%	s 3*	309
so 4*	253	91%	s 4*	24
so 5*	5	2%	s 5*	218
so 6*	132	79%	s 6*	36

so 7*	118	92%	s 7*	10
so 8*	1	1%	s 8*	125
so 9*	2		s 9*	85

Table 50. Vocalisation of the preposition **s** in the Slovak corpus.

Judging from the data in Table 50, it becomes quite clear that the vocalised form of the preposition **s** is used practically obligatorily before words beginning with the sibilants ⟨s⟩, ⟨š⟩, ⟨z⟩, ⟨ž⟩.

so s*	22,846	100%	s s*	36
so š*	3,796	99%	s š*	21
so z*	10,403	100%	s z*	9
so ž*	2,225	99%	s ž*	30
Total	39,270	100%	Total	96

Table 51. Vocalisation of the preposition **s** before words beginning with a sibilant consonant.

Out of the 96 instances of **s** before words beginning with a sibilant, the majority, or as many as 62 instances, occur in the older texts of the corpus, which might be interpreted as a sign of a slightly unstable situation a century ago. It is interesting to note that the collocation **s sebou** *with oneself* occurs 16 times in the older texts, but not even once in the contemporary texts.

Only 34 instances of **s** before words beginning with a sibilant occur in contemporary texts. Out of these occurrences, 11 instances occur in death notices where a name beginning with a sibilant is printed on a separate line and the preposition is found on the preceding line, e.g. **Zajtra (v piatok 2.8) sa navždy rozlúčime // v košickom krematóriu s // Štefanom GALOM // 62-ročným // o 12.00 hod.** *Tomorrow (on Friday 2.8) we will forever say farewell // in the Košice crematorium to // Štefan GAL // 62 years old // at 12 noon.* This leaves us with only 23 contemporary instances, all of very individual character, and not exhibiting any noticeable pattern, such as e.g. **s životnými skúsenosťami** *with experiences from life*. They constitute less than 0.1% and could therefore be interpreted as typing mistakes or possibly deliberate deviations by the individual writer from a very clear principle.

The only other contexts that trigger more frequent vocalisation of the preposition **s(o)** are words beginning with ⟨m⟩ and ⟨v⟩. These contexts will therefore be investigated more closely. The relevant data is presented in the following two tables.

so ministerkou	1		s ministerkou	107
so mňa	13		s mňa	0
so mnou	1,976		s mnou	0

so mzdami	13		s mzdami	2
so mzdou	4		s mzdou	0
so mzdovou	4		s mzdovou	0
so mzdovým	3		s mzdovým	0
so mzdovými	5		s mzdovými	0

Table 52. Vocalisation of the preposition *s* before words beginning with ⟨m⟩.

The context words beginning with ⟨m⟩ are restricted to four lexemes, of which **so ministerkou** *with the minister* occurs only once and is best interpreted as a writing mistake, as there are 107 instances of the contrasting collocation with the non-vocalised form of the preposition. Before the 1st person singular personal pronoun, only the vocalised form of the preposition occurs, and before the words **mzda** *wages*, as well as the corresponding relational adjective **mzdový**, there are 29 instances of the vocalised form and only two instances of the unvocalised form of the preposition. The relatively small number of occurrences of **so mňa** is explained by the fact that this spelling is outdated, and the modern Slovak orthography prescribes **zo mňa**, as **s(o)** should only be used with the instrumental case, whereas **z(o)** is used with the genitive case. In older orthography, **s(o)** with the genitive was used to denote movement from a surface, as opposed to **z(o)** with the genitive denoting a movement from the inside of an object.

After a look at the data in contexts with a following word beginning with ⟨v⟩, it immediately becomes evident that the following letter is significant. Once again, the sibilants constitute a triggering factor.

In this specific context, a few new instances of misleading data were located and disregarded, namely transcribed Russian text, the German **so**, the misspelt **so vojimi**, **so vojou** instead of **so svojimi**, **so svojou**, as well as **so VÚB**, which stands for **so Všeobecnou úverovou bankou**. Such misleading data have been omitted in the following table.

so vč*	12		s vč*	14
so vd*	1		s vd*	10
so vch*	1		s vch*	3
so vs*	127	100%	s vs*	0
so vš*	1667	99%	s vš*	11
so vt*	1	1%	s vt*	102
so vz*	340	99%	s vz*	2
so vž*	3		s vž*	1

Table 53. Vocalisation of the preposition *s* before words beginning with ⟨v⟩ followed by a consonant.

Out of the 14 instances with **s vš***, **s vz*** and **s vž***, only one instance in each of

these three contexts occurs in contemporary text. It therefore seems that the usage of the vocalised form of the preposition is very close to obligatory before consonant clusters consisting of /v/ followed by a sibilant fricative. In the context before /vč/, both forms of the preposition seem to be equally frequent.

Before all remaining initial letters the frequency of vocalisation is below 1%. All non-individual instances are listed below and compared to the occurrences of the contrasting collocation with the non-vocalised form of the preposition. As there are many numerals in this remaining material, the numerals are discussed separately.

so ctihodným	2		s ctihodným	1
so cfou	37		s cfou	0
so dňa	80		s dňa	2
so džbáňkom	2		s džbáňkom	0
			s džbáňom	4
so lžou	3		s lžou	0
so ps*	161	76%	s ps*	51
so pseudo*	2		s pseudo*	5
so psych*	16		s psych*	43
so pšen*	8		s pšen*	6
so pštroš*	3		s pštroš*	4

Table 54. Vocalisation of the preposition **s** before words beginning with consonants other than sibilants, ⟨m⟩ or ⟨v⟩.

In the collocational type **so cfou** *with honour*, the vocalised form of the preposition appears to be obligatory.

All instances of **so** before ⟨d⟩ occur in older texts. The spelling **so dňa** *from a/the day*, containing **s(o)** followed by the genitive case, is outdated and is replaced by **zo dňa** in contemporary Slovak. It should also be mentioned that the vast majority of these occurrences constitute part of the longer collocation **so dňa na deň** *from one day to another*.

Before the root {lž} *lie*, no contrasting occurrences with the non-vocalised form of the preposition are found in the corpus texts. The vocalised form is found in the collocations **so lžou** (3/0), **so lžami** (1/0), **so lživým** (1/0).

Again, in the context before a following word beginning with ⟨p⟩ followed by a sibilant, a few abbreviations have been disregarded, such as **s PSL**. Lexemes which are based on the root {ps} *dog* exhibit 141 occurrences in collocations with the vocalised form of the preposition, as opposed to only 3 occurrences with the non-vocalised form. All these three occurrences are found in the older texts. It therefore seems that the vocalised form is obligatory before all lexemes formed from the root {ps}, e.g. **so psami** *with the dog*, **so psíkom** *with the little dog*, **so psím mäsom** *with dog meat*.

Only one instance each is at hand in the corpus of **so psoriázou** and **so pstruhmi**. These collocations might imply the possibility of a general tendency to vocalise before ⟨p⟩ followed by a sibilant.

Before two much more frequent morphemes of foreign origin, ⟨psych⟩ and ⟨pseudo⟩, the vocalised form of the preposition is less frequent. Here, the non-vocalised form is about twice as frequent as the vocalised form.

Finally, before ⟨pš⟩ both forms of the preposition occur. There is a strong tendency towards the vocalised form in the older texts, whereas there is an equally strong tendency towards the non-vocalised form in the contemporary texts.

At this point only the numerals remain to be investigated. All non-unique occurrences of collocations with the vocalised form of the preposition are presented in the following table.

so dva	8	s dva	1
so dve	4	s dve	0
so dvoch	5	s dvoch	9
so dvoma	2	s dvoma	730
so pár (+ gen.)	21	s pár (+ ins.)	33
so päť (+ gen.)	2	s päť a pol (+ ins.)	2
so traja	2	s traja	0
so tri	12	s tri	0
so troch	4	s troch	0

Table 55. Vocalisation of the preposition **s** before numerals.

Once again, it should be pointed out that in the contemporary orthography, **s(o)** is only used with the instrumental case in the meaning *with*. In none of the collocations with the vocalised form in the above table, is the numeral in the instrumental. Instead, there are instances of the nominative, accusative and/or genitive. Also all the instances of **s dva** and **s dvoch** are found in the older texts. The usage with the genitive in the meaning *of, from* can readily be explained by the older orthographical principles, but instances with the nominative or accusative cannot. Instead, the answer is to be found in the semantics of **so** in these collocations, where it has the function of an approximative adverbial particle meaning *about* or *some*. The contemporary orthography prescribes the spelling **zo** in this function, which in KSSJ'97 is found as an independent lexeme without any shorter “non-vocalised” form. Therefore, we will return to this discussion in connection with the preposition **z(o)**.

As can be seen in Table 50 above, the same pattern occurs before digits, as has already been shown to occur in the rest of the material, namely a very high frequency of vocalisation before 4, 6, 7, which begin with a sibilant, as well as

medium frequency before 1, which may well be explained by the fact that certain numbers such as 14, 16, 17, 100 begin with a sibilant, whereas many others beginning with the same digit do not.

However, as there is no way to know with certainty how the digits would be read out by the writer, these data will not be investigated further. Slovak, similarly to Czech, has the possibility of inverting the order of the ones and the tens, although it is not as frequent in Slovak as in Czech, so that 72 can be read as **sedemdesiatdva** or **dvaasedemdesiat**. Also, the writer's awareness of how the digits may influence the choice of preposition may be lower when the writer does not pronounce the text to himself.

7.1.2.3. v(o)

All the instances of the form **vo** were extracted from the corpus. Some occurrences were found to be instances of printing mistakes where words had been divided into two parts, such as **vo lby** instead of the correct **voľby elections**. A few typing mistakes in the word following **vo** also occurred, e.g. **vo nme** instead of the correct **vo mne in me**. Furthermore, vocalisation occurred about 90 times before words with an initial consonant group not complying to the principles of Slovak orthography, such as **vo Phoenixe**, **vo vo Schwerine**, **vo squashy**, **vo Zwickau**. The vocalised form **vo** also occurred in 162 instances before the Roman ordinal number **IV**, most probably pronounced as **vo štvrtom** (masculine and neuter) or **vo štvrtej** (feminine). Finally, it occurred once in a prayer in the collocation **vo iskušeníje**, which should be interpreted as an allusion to Church Slavonic.

Contrasting data with the non-vocalised form of the preposition were also retrieved from the corpus in all the contexts in which the vocalised form was encountered. Here as well, some misleading data were localised, such as a few printing mistakes where **v** occurs instead of **s** (4 times) and **z** (1 time). There were also a dozen instances of a single **V** being an initial, over 20 instances in which it was a Roman numeral, and a handful of pure spelling mistakes, such as, for instance, **v včeskoslovensku in Czechoslovakia**. There were also three occurrences in Czech.

The non-vocalised form of the preposition also occurred over 30 times before various abbreviations beginning with **F**, which as a rule is read out so that the initial letter is pronounced [ef], hence e.g. **v FK** is pronounced as [vɛf'ka:]. Finally, there were also 25 instances where the non-vocalised form of the preposition occurred before ordinal numbers expressed as Roman numerals beginning with a **V**, e.g. **v VI. ročníku** meaning **v šiestom ročníku in the sixth grade**. All such misleading data have been omitted in the following table.

vo b*	9	0%	v b*	37,694
vo c*	420	4%	v c*	11,381
vo č*	2	0%	v č*	14,154
vo d*	1,593	5%	v d*	33,133
vo f*	9,763	100%	v f*	8
vo h*	88	0%	v h*	19,297
vo j*	5	0%	v j*	18,428
vo k*	21	0%	v k*	55,008
vo l*	6	0%	v l*	13,060
vo l̄*	8	1%	v l̄*	1,220
vo m*	771	2%	v m*	33,921
vo n*	5	0%	v n*	53,000
vo o*	1	0%	v o*	29,695
vo ô*	1	1%	v ô*	130
vo p*	22	0%	v p*	97,707
vo r*	6	0%	v r*	52,153
vo s*	11,299	13%	v s*	77,051
vo š*	4,533	28%	v š*	11,525
vo t*	124	0%	v t*	68,605
vo t̄*	1	0%	v t̄*	543
vo ú*	1	0%	v ú*	7,705
vo v*	38,548	100%	v v*	34
vo w*	1,273	100%	v w*	2
vo z*	1,379	5%	v z*	25,167
vo ž*	5	0%	v ž*	6,004
vo 2*	2	0%	v 2*	1581
vo 4*	50	7%	v 4*	643

Table 56. Vocalisation of the preposition **v** in the Slovak corpus.

As could be expected from the background on the previous two asyllabic prepositions in Slovak, there is practically total vocalisation at hand before the graphemes ⟨v⟩, ⟨w⟩, and ⟨f⟩, i.e. before the phonemes /v/ and /f/.

Of the eight instances where vocalisation is lacking before ⟨f⟩, two instances occur in poetry within the older texts. The rest are contemporary, mainly from newspaper text, and cannot be explained other than as writing mistakes or as individual choices. Four out of the six contemporary instances involve proper names, e.g. **v Filipínskej metropole Manile** in the *Philippine metropolis Manila*. Before ⟨w⟩, both occurrences are contemporary and involve proper names, e.g. **v Warnemunde** in *Warnemünde*. Finally, before ⟨v⟩ only 12 of the 34 occurrences with the non-vocalised form of the preposition occur in contemporary newspaper text, whereas the remaining 22 instances occur in the older part of the corpus and 15 of them are found in poetry. Once again,

this indicates a certain amount of instability in the older stages of Slovak and a certain degree of individual freedom which manifests itself above all in poetry. It should also be noted that out of the 22 instances in the older texts, 4 were found in texts by Jaroslav Vlček and 12 in texts by Ján Kovalík Ústiansky, which indicates that such vocalisations are to a large extent based on the individual preferences of certain individuals. When it comes to the contemporary occurrences, no pattern is evident, and considering the marginal number of such occurrences, they are probably best interpreted as typing mistakes. In contemporary Slovak, vocalisation of the preposition **v(o)** could therefore be considered to be obligatory before the phonemes /f/ and /v/.

Another context which once again generates many vocalisations is that of sibilants. A closer analysis was therefore carried out based on the grapheme immediately following the initial sibilant. Contexts which only generated one single instance have been excluded from the table. Such contexts mainly consisted of a sibilant followed by a vowel, and these instances are most probably typing mistakes, as vocalisation is highly implausible in that type of context. The data concerning the sibilant initial contexts are presented in the following table.

vo sf*	71	71%	v sf*	29
vo sk*	33	1%	v sk*	4,143
vo sl*	15	0%	v sl*	6,539
vo sm*	12	1%	v sm*	816
vo sn*	294	19%	v sn*	1,219
vo sp*	49	1%	v sp*	7,896
vo sr	4	0%	v sr*	1,472
vo st*	22	0%	v st*	11,663
vo sv*	10,711	86%	v sv*	1,777
vo šk*	7	0%	v šk*	1,974
vo šm*	2		v šm*	9
vo šp*	6	0%	v šp*	1,625
vo št*	3,508	43%	v št*	4,679
vo štv*	3,502	93%	v štv*	283
vo šv*	1,002	95%	v šv*	51
vo zb*	2	0%	v zb*	504
vo zd*	7	1%	v zd*	884
vo zh*	5	3%	v zh*	182
vo zm*	13	1%	v zm*	1,722
vo zr*	3	6%	v zr*	461
vo zv*	1,345	90%	v zv*	143
vo ži*	3	0%	v ži*	4,597

Table 57. Vocalisation of the preposition **v** before sibilants.

A closer look at these data suggest that the majority of instances with the vocalised form of the preposition, followed by a sibilant but not followed by ⟨v⟩, are found in the older texts of the corpus. A revised form of the data that excludes the older texts gives a much clearer picture of the distribution. These contemporary data are presented in the table below, sorted according to relative frequency.

vo žv* ⁷⁷	0		v žv*	0
vo sv*	8,706	97%	v sv*	246
vo šv*	980	97%	v šv*	33
vo sn*	194			
vo sne	194	97%	v sn*	7
vo zv*	1,141	94%	v zv*	69
vo št*	3,438			
vo štv*	3,438	94%	v štv*	228
vo sf*	70		v sf*	22
vo sk*	8			
vo skv*	8		v skv*	44
vo sm*	1	0%	v sm*	501
vo sp*	2	0%	v sp*	7,106
vo st*	2	0%	v st*	10,045
vo šp*	2	0%	v šp*	1,518
vo zm*	2	0%	v zm*	1,586

Table 58. Vocalisation of the preposition *v* before sibilants in the contemporary part of the Slovak corpus.

The data in Table 58 show less variation in vocalisation in the contemporary part of the corpus, but there is still a very strong tendency to use the vocalised form of the preposition *v(o)* before words beginning with a sibilant immediately followed by a ⟨v⟩, and a somewhat weaker, but still quite strong, tendency before words beginning with a sibilant immediately followed by an ⟨f⟩. One plausible reason for this weaker tendency before sibilant + ⟨f⟩ is that this type of cluster does not occur in native Slovak words, but only in foreign borrowings. There is also a very strong tendency to vocalise this preposition before the consonant cluster ⟨štv⟩, as well as a weak tendency to vocalisation before the cluster ⟨skv⟩.

⁷⁷ The context ⟨žv⟩ has been added to the table in order to show that there are no instances of any of the two forms of the preposition followed by a word beginning with this consonant cluster.

All the collocations with **v(o)** followed by ⟨štv⟩ or ⟨skv⟩ include one of the stems in the following table. In the contexts of these two initial consonant clusters, collocational types containing three different stems prove to be quite frequent and exhibit a certain degree of vocalisation, whereas three other collocational types do not exhibit any vocalisation at all.

vo skvalit*	0		v skvalit*	6
vo skvel*	8		v skvel*	36
vo skvost*	0		v skvost*	2
vo štvav*	0		v štvav*	3
vo štvor*	171	64%	v štvor*	97
vo štvrť*/štvrť*	3267	96%	v štvrť*/štvrť*	128

Table 59. Vocalisation of the preposition **v** before ⟨štv⟩ and ⟨skv⟩ in the contemporary part of the Slovak corpus.

From the above data it also follows that the relative frequency of the vocalisation of **v(o)** increases with the frequency of the individual collocational type. In the table below, a closer investigation of the most frequent collocational types reinforces this observation even further.

vo štvrtok	2682	100%	v štvrtok	4
vo štvrtkov*	59		v štvrtkov*	2
vo štvrťfinále	330	95%	v štvrťfinále	19
vo štvrtom / štvrtej / štvrtých	157	67%	v štvrtom / štvrtej / štvrtých	76
vo štvrti	16	43%	v štvrti	21

Table 60. Vocalisation of the preposition **v** before ⟨štv⟩ in the contemporary part of the corpus.

When it comes to the remaining contexts triggering vocalisation of the preposition **v(o)**, the occurrences are fewer. All such contexts with more than one occurrence are listed in the following table.

vo dne	73		v dne	0
vo dverách	53		v dverách	2
vo dvoch	71	8%	v dvoch	844
vo dvoj*	63	16%	v dvoj*	321
vo dvor*	57		v dvor*	23
vo hvezdární	2		v hvezdární	2
vo hviezd*	66		v hviezd*	17
vo chvál*	2		v chvál*	3
vo chvíľ*/chvíľ*	284	99%	v chvíľ*/chvíľ*	4

vo chvo*	3	v chvo*	0	
vo Lvove	6	v Lvove	3	
vo mne	195	100%	v mne	0
vo mzd*	8	v mzd*	70	
vo Rwande	2	v Rwande	17	
		v rwandskej	1	

Table 61. Vocalisation of the preposition **v** in remaining contexts in the contemporary part of the Slovak corpus.

An analysis of the data presented in the above table suggests that vocalisation of the preposition **v(o)** is obligatory in the collocations **vo dne** (73/0) *in the day*, **vo dvore** (50/0) *in the yard*, **vo hviezdach** (44/0) *in the stars*, **vo mne** (195/0) *in me*. The singular **vo hviezde** (1/0) *in the star*, only occurring once, presumably belongs to the same category. Other collocational types only occurring once or twice are **vo chválach** (2/0) *in the praises*, **vo Chvojnici** (2/0) *in Chvojnica*, **vo chvoste** (1/0) *in the tail*, and therefore their status is quite uncertain.

Furthermore, the vocalised form is highly frequent, but not obligatory, in **vo dverách** (53/2) *in the door*, **vo dvojici** (25/7) *in the pair*, **vo chvíli** (211/1) *in the moment*, **vo chvíľach** (69/1) *in the moments*, **vo Lvove** (6/3) *in Lviv*, **vo mzde** (8/2) *in the salary*. Because there are so few instances, the diminutives **vo chvíľke** (4/0) *in the little moment* and **v chvíľkach** (1/0) *in the little moments*, exhibiting vocalisation in the singular, but not in the plural, will be presumed to belong to the same category, whereas the single instance of the relative adjective **v chvíľkovom** (1/0) might be interpreted as an effect of a general weaker tendency to vocalise before adjectives than before nouns.

In derivational forms of the noun **hviezda** *star*, mainly before relational adjectives, vocalisation is equally as frequent as non-vocalisation, e.g. **vo hviezdnom** (14/15), **vo hviezdnej** (3/0), **vo hviezdnych** (1/2), but this also seems to be true for the derived noun **vo hvezdárni** (2/2) *in the observatory*.

Finally, vocalisation is possible, but not very frequent, in the collocational types **vo dvoch** (71/844) *in two*, **vo dvojhách** (13/30) *in the singles*, **vo dvojhre** (18/52) *in the single*, **vo dvojiciach** (3/10) *in pairs*, **vo dvorane** (2/1) *in the hall*, **vo dvoroch** (2/1) *in the yards*, as well as with the proper geographical names **vo Dvoroch nad Žitavou** (2/13), **vo Dvore Králové** (1/3), **vo Rwande** (2/17).

Interestingly enough, the absolute majority of the words following the vocalised preposition **vo** begin with a consonant cluster containing /v/. However, if this /v/ is non-initial and the initial consonant is not a sibilant, there seems to be no tangible systemic tendency to vocalise before such clusters, but rather some sort of lexicalisation of a certain smaller set of collocational types. Such infrequent lexicalisations may then constitute part of the lexicon only for some speakers and not for others.

7.1.2.4. z(o)

All the occurrences of the forms **z** and **zo** were extracted from the corpus and the data were sorted and counted according to the initial grapheme of the following word. In this process, a number of misleading data were encountered.

There were about 200 instances of the abbreviation **ZO**. There were also two instances of the vocalised form of the preposition before abbreviations, i.e. **zo FC Sao Paolo**, **zo CME**, where it is unclear how these abbreviations would be pronounced by the author of these texts. There were also some instances of the vocalised form of the preposition before foreign names not complying to the principles of Slovak orthography, such as **zo Cindy**, **zo Champs Elysées**. Furthermore, there were more than 20 instances of the vocalised preposition before Roman numerals representing ordinal numbers, such as **zo XVI**. Finally, there were over 60 typing mistakes such as **zo Ľivota** instead of the correct **zo života** *from life*, as well as a few instances of **zo** instead of the correct **za** *for* or **to** *that*.

The non-vocalised form **z** was also found in misleading occurrences not representing the preposition investigated. There were 7 instances of the letter of the alphabet and 4 occurrences as an initial. Further, there were over 300 instances of the preposition **z** before a following abbreviation beginning with **S** or **Š**, such as **z SDĽ**, but all such abbreviations are generally pronounced with an initial [ɛs] or [ɛʃ]. There were also a few instances of the non-vocalised form of the preposition before foreign names not complying with the principles of Slovak orthography, such as **z Zwittau**, and there were six instances of **z** in linguistic metatext discussing the forms **z/zo**. There were also a few writing mistakes, such as **z spadla nôh** or **z spomedzi** instead of the correct **spadla z nôh** *fell from her feet* and **spomedzi** *from among*, as well as two instances in old Czech text and three instances in modern Czech.

In the following table, all such misleading data have been disregarded.

zo a*	2	0%	z a*	3,743
zo b*	6	0%	z b*	7,876
zo c*	1	0%	z c*	5,620
zo č*	1	0%	z č*	5,101
zo d*	1,015	7%	z d*	13,733
zo f*	1	0%	z f*	3,847
zo j*	1	0%	z j*	6,295
zo l*	5	0%	z l*	4,621
zo m*	407	3%	z m*	11,630
zo p*	101	0%	z p*	25,882
zo r*	1	0%	z r*	11,734
zo s*	23,316	100%	z s*	13

zo š*	4,768	100%	z š*	11
zo t*	116	0%	z t*	24,013
zo v*	2,557	16%	z v*	13,404
zo x*	3		z x*	0
zo z*	9,751	100%	z z*	8
zo ž*	1,778	99%	z ž*	17
zo 1*	911	26%	z 1*	2527
zo 2*	12	1%	z 2*	1372
zo 3*	20	3%	z 3*	754
zo 4*	1,522	99%	z 4*	15
zo 5*	2	0%	z 5*	399
zo 6*	409	97%	z 6*	15
zo 7*	285	99%	z 7*	3
zo 9*	1	0%	z 9*	232

Table 62. Vocalisation of the preposition *z* in the Slovak corpus.

From the data in the above table, it immediately becomes clear that the highest frequency of vocalisation occurs before words beginning with one of the sibilants *s*, *š*, *z*, *ž*. The data in these four contexts are summarised in the following table.

zo S*	39,613	100%	z S*	49
-------	--------	------	------	----

Table 63. Vocalisation of the preposition *z* before sibilants in the Slovak corpus.

Out of the 49 instances with the non-vocalised form of the preposition, only 18 occur in the contemporary part of the corpus, and they all consist of single occurrences of individual collocations. The vocalised form of this preposition can therefore be considered obligatory in the context before words beginning with a sibilant consonant, and only some very marginal individual departures from this principle are encountered. Such departures are likely to be typing mistakes.

For the other contexts with a relatively high frequency of the vocalised form *zo*, it turns out that 54% of occurrences before ⟨d⟩, 56% before ⟨m⟩ and 37% before ⟨v⟩ are encountered in the older part of the corpus. Therefore a new search was made exclusively in the contemporary texts, giving the following results.

zo desať	9	56%	z desať	7
zo dna	3	25%	z dna	9
zo dňa	424	99%	z dňa	4
zo dva	11	100%	z dva	0
zo dvadsať	4	80%	z dvadsať	1

zo dvaja	3	100%	z dvaja	0
zo dve	9	100%	z dve	0
zo mňa	117	100%	z mňa	0
zo mzdových	3	75%	z mzdových	1
zo mzdy	61	100%	z mzdy	0
zo vč*	87	81%	z vč*	20
zo včera	5	100%	z včera	0
zo včerajšej / -ieho /- ích	63	84%	z včerajšej / -ieho /- ích	12
zo včerajška	18	100%	z včerajška	0
zo vs*	59	100%	z vs*	0
zo Všetína	12	100%	z Všetína	0
zo vstup*	45	100%	z vstup*	0
zo vše*	1,205	>99.9%	z vše*	1
zo vzácn*	8	89%	z vzácn*	1
zo vzájomn*	51	100%	z vzájomn*	0
zo vzdel*	11	100%	z vzdel*	0
zo vzdial*	35	100%	z vzdial*	0
zo vzdoru	3	100%	z vzdoru	0
zo vzdu*	90	99%	z vzdu*	1
zo vzkriesenie	2	100%	z vzkriesenie	0
zo vznešen*	5	100%	z vznešen*	0
zo vznik*	8	89%	z vznik*	1
zo vzor*	14	100%	z vzor*	0
zo vzrast*	6	86%	z vzrast*	1
zo vzťah*	14	100%	z vzťah*	0

Table 64. Vocalisation of the preposition *z* before ⟨d⟩, ⟨v⟩, and ⟨m⟩ in the contemporary part of the Slovak corpus.

A dozen single occurrences such as **zo daného štátu** *from the given state of* are excluded from the above table as they do not reveal any systematic patterns.

Of the four collocational types occurring before words beginning with ⟨d⟩, **zo dna** *from the bottom* appears to be a possible, but less frequent variant, whereas the vocalisation in **zo dňa** *from the day* appears to be very close to obligatory. The collocations with the numerals **desať** *ten* and **dva** *two* and their derivational forms deserve more detailed comment, as there are two completely different meanings of **zo** at hand. The preposition *z* with its standard meaning *from, of* governs the genitive case, and there are many such occurrences with the non-vocalised form of the preposition in the corpus, i.e. **z desiatich** *of ten*, **z dvoch** *of two*. In some contexts, however, there is also the possibility of leaving the numeral undeclined, which generates e.g. **rozšíriť z desať na dvanásť tímov** *to extend from ten to twelve teams* and **z dvadsať**

výstrelov *out of 20 shots*. However, in the above table all the occurrences of **zo** followed by a numeral contain a semantically separate lexeme constituting an adverbial particle which expresses approximation, e.g. **zo desať** *about ten*.

When it comes to the collocations with the vocalised preposition before words beginning with ⟨m⟩, the vocalisation before the 1st person singular personal pronoun, **zo mňa** *from me*, is evidently obligatory, as could be expected on the basis of this pronoun's behaviour in relation to other prepositions. The vocalised form also seems to be obligatory in the collocation **zo mzdy** *from the salary*, whereas some variation is at hand in the scanty data containing collocations with the corresponding relative adjective **mzdový**.

When it comes to collocations with the vocalised form **zo** before words beginning with ⟨v⟩, the data clearly shows that vocalisation is practically obligatory before words beginning with ⟨v⟩ followed by a sibilant.

zo vS*	1516	100%	z vS*	5
--------	------	------	-------	---

Table 65. Vocalisation of the preposition **z** before ⟨v⟩ followed by a sibilant in the Slovak corpus.

The vocalised form **zo** is also very frequent but not obligatory before words beginning with ⟨vč⟩. The tendency is strongest in the collocations **zo včera** *from yesterday* (5/0) and **zo včerajška** *from yesterday* (18/0), and somewhat weaker in collocations with the corresponding relative adjective **zo včerajšieho/včerajšej/včerajších** *from yesterday's* (63/12). It should particularly be noted that vocalisation occurs only before the root {včer} *yesterday*, not before other roots, e.g. {včel} *bee*.

The rest of the collocations with the vocalised form **zo** in the contemporary part of the corpus are given in the table below.

zo bz*	3	z bz*	0
zo lž*	6	z lž*	0
zo pä*	13	z pä*	130
zo ps*	23	z ps*	51
zo pš*	2	z pš*	8
zo tr*	22	z tr*	2,880
zo x*	3	z x*	0

Table 66. Vocalisation of the preposition **z** in infrequent contexts.

The following single instances are not included in the above table: **zo a do zamestnania** *from and to work*, **zo aj do zahraničia** *from and to abroad*, **zo bavineného** *of cotton*, **zo bohoslužbami** *with the church services*, **zo finále** *from the final*. The first two instances are quite interesting as they show that vocalisation may serve as a strategy in order to pronounce a stressed

preposition as an independent word. The following two instances, however, seem to be mere typing mistakes, as no tendency has been registered elsewhere to vocalise prepositions before a non-homorganic consonant followed by a vowel.

It can be observed that contexts with a consonant cluster including a sibilant exhibit a higher frequency of vocalisations than clusters without a sibilant, which once more confirms previous observations. The three instances of **zo bz*** are **zo Bzenca**, **zo Bzenova**, **zo Bzineckých kopańíc**. The six instances of **zo lz*** are **zo lži** (5/0) and **zo lživej** (1/0). The 23 instances of **zo ps*** (12/0) are different forms of the root meaning *dog*, **zo pseudo*** (1/3) and **zo psych*** (10/48). The two instances of **zo pš*** are **zo pštros*** (2/3) of *ostrich*, and it should be especially noted that there are no instances of the vocalised form in the collocational type **z pšeni*** (5/0) of *wheat*. The three instances of **zo x*** are **zo xenofóbie** (2/0) and **zo Xibalby** (1/0). The relatively low frequency before ⟨ps⟩ may partly be interpreted as a lesser tendency to vocalise before words of foreign origin than before native Slovak words. The low frequency before /pš/ is much more difficult to explain, but on the other hand, the amount of data is relatively small.

The contexts not including a sibilant are represented exclusively by numerals. The 13 instances of **zo pä*** are **zo päť** (11/11), **zo päťnásť** (1/0) and **zo päťtisíc** (1/1), whereas the 22 instances of **zo tr*** are **zo tri** (16/1), **zo tridsať** (3/0), **zo trikrát** (3/0) and **zo troch** (1/536). All these collocations involve instances of the approximative adverbial particle **zo** *about* and none of the forms following the preposition is in the genitive case, which is required by the preposition **z(o)** *of, from*, as the form **troch** *three* in the collocation **zo troch** is actually in the accusative case, formally equal to the genitive in the masculine plural paradigm, and **zo** is once more an instance of the approximative adverbial particle: **stihol som ešte zo troch konkurentov predbehnúť** *I still managed to outstrip about three more competitors*.

Finally, a short comment on the collocations involving digits which are found at the bottom of Table 62. These data will not be analysed in any detail, as it is not possible to know exactly how the digits would be read out by the individual authors. It is quite clear, however, that these data confirm the general tendencies observed elsewhere in the material, i.e. vocalisation before words beginning with a sibilant, e.g. **štyri** *four*, **šesť** *six*, **sedem** *seven* and **sto** *hundred*.

7.2. The Literature

There is somewhat less written on the issue under scrutiny with regard to Slovak than with regard to Polish. Also, the publication of codificational works, linguistic handbooks and monolingual Slovak dictionaries has traditionally been concentrated in the two publishing houses VEDA,

associated with the Academy of Sciences, and ŠPN, specialising in pedagogical publications.

7.2.1. Dictionaries

The Slovak language has been described in noticeably fewer dictionaries than Polish. There are basically three modern dictionaries. *Slovník slovenského jazyka* (SSJ'59) was published in six volumes in the years 1959–1968. *Krátky slovník slovenského jazyka* (KSSJ'97) was published in three different editions: 1987, 1989 and 1997, and was produced as a condensed version of SSJ'59. The new *Slovník súčasného slovenského jazyka* (SSSJ'06) is based on the Slovak National Corpus. It has so far appeared in two volumes (2006, 2011) and has reached the letter L.

SSJ'59 comments on the usage of the vocalised shape of some of the prepositions, but not all. KSSJ'97 contains even less information on the usage of the vocalised forms. SSSJ'06 has not yet reached the more interesting letters of the alphabet with regard to the vocalisation of prepositions.

Besides the explanatory dictionaries, Slovakia has a long tradition of orthographic dictionaries called *Pravidlá slovenského pravopisu* (PSP), which are published in new editions at shorter intervals than the others. Some of the editions also contain quite extensive grammatical sections, mainly concerned with detailed principles for Slovak spelling in quite a broad sense, and in some editions the vocalisation of prepositions has also been discussed in these sections. Such descriptions will be reviewed, together with the handbooks on Slovak grammar.

All the dictionaries mentioned, as well as a few others, are conveniently searchable on the Internet at <<http://slovniky.korpus.sk/>>.

7.2.1.1. bezo

SSJ'59 and SSSJ'06 say that this form is only used in the collocation **bezo mňa**. The statements in KSSJ'97, PSP'00 and PSP'05 are less categorical, simply saying that this form is used in the collocation **bezo mňa**.

In the corpus examined, a number of other collocations occurred with significant frequencies. Therefore, for comparison, the corresponding statistical data were retrieved from the Slovak National Corpus⁷⁸:

bezo dna	6	6%	bez dna	96
bezo mňa	1,154	100%	bez mňa	4
bezo slov	7	1%	bez slov	1,160
bezo slova	1	0%	bez slova	3,128

⁷⁸ Slovenský národný korpus at <<http://korpus.juls.savba.sk/>>.

bezo sna	2		bez sna	30
bezo sporu	23		bez sporu	17
bezo zbytku	41		bez zbytku	6
bezo zmeny	335	17%	bez zmeny	1,644
bezo zmien	131	18%	bez zmien	599
bezo zvyšku	476	60%	bez zvyšku	312

Table 67. SNK data as of 14 August 2012 regarding collocations with **bezo**.

The SNK data confirm the prescriptive standpoint of the codificational literature with a rate of vocalisation before the pronoun **mňa** that is practically 100%.

The collocational type **bezo zvyšku** *completely, without remainder* dominates slightly over the non-vocalised form in SNK, but the relative frequency of the remaining collocational types with the vocalised form of the preposition is rather low. With regard to **bezo sporu** *without dispute*, **bezo zbytku** *completely, without remainder* and **bezo sna** *without sleep*, the absolute frequency is also low, and therefore their frequencies have been retrieved with Google⁷⁹: **bezo sporu** 51% (5,670/5,330), **bezo zbytku** 41% (10,300/14,800) and **bezo sna** 0% (8/9,790). **Bezo zmien** *without changes* and **bezo zmeny** *without change* are evidently less frequent than the corresponding collocation with the non-vocalised form. **Bezo dna** *without bottom*, **bezo slov** *without any words*, **bezo slova** *without a word*, together with **bezo sna**, are rare. With the exception of a few collocations with a very low absolute frequency, the previously presented corpus findings are well in line with the SNK data.

7.2.1.2. cezo

SSJ'59 and SSSJ'06 state that the form **cezo** is only used in the collocation **cezo mňa**, whereas KSSJ'97, PSP'00 and PSP'05 say only that this form is used in the collocation **cezo mňa**.

For comparison, the corresponding data retrieved from SNK are presented in the following table.

cezo mňa	420	95%	cez mňa	22
----------	-----	-----	---------	----

Table 68. SNK data as of 14 August 2012 regarding collocations with **cezo**.

⁷⁹ All Slovak data emerging from Google were retrieved in August 2012 using the search syntax "**search string** **site:sk**", where the search string is placed within citation marks in order to generate a more precise search, and **site:sk** generates a search restricted only to pages with an address ending in <.sk>. The exactness of the searches is however somewhat problematic, as intervening punctuation marks are disregarded by Google.

The dictionaries as well as the SNK data confirm the previously discussed corpus findings. Individual variation is evidently somewhat larger with this less common preposition, but the relative rate is still very high.

7.2.1.3. **nado**

SSJ'59 says that the form **nado** is only used in the collocations **nado mňa**, **nado mnou**, **nado dvermi**. KSSJ'97 says it is used in the collocations **nado mňa**, **nado mnou**, **nado dvere**, **nado dvermi**. PSP'00 and PSP'05 do not list any collocations for this preposition.

In the corpus investigated, a number of other collocations also occurred with a significant frequency of vocalisation. For comparison, the corresponding data retrieved from SNK are therefore presented in the following table.

nado dňom	5		nad dňom	23
nado dverami	2	1%	nad dverami	264
nado dvere	4		nad dvere	43
nado dvermi	26		nad dvermi	31
nado mňa	110	99%	nad mňa	1
nado mnou	2,215	100%	nad mnou	11
nado všetko	84	19%	nad všetko	348
nadovšetko	3,182			
nado všetky	22	4%	nad všetky	579
nado všetkých	0	0%	nad všetkých	112

Table 69. SNK data as of 14 August 2012 regarding collocations with **nado**.

The independent data confirm each other to a very high degree. The vocalisation is practically obligatory before the first person personal pronoun. The vocalised form also occurs at a significant rate in the collocation **nado všetko** *above all*, where it is mostly used in an adverbial meaning. This lexicalised adverbial phrase should however be written as one word according to the contemporary Slovak orthography. In all other contexts the vocalised form is rare, with the exception of the collocational type **nado dvermi** *above the door* in SNK.

As the absolute frequency of some collocations is low, their frequencies have been retrieved with Google: **nado dňom** *towards daybreak* 9% (1,350/13,200), **nado dverami** 0% (49/37,000), **nado dvere** 0% (218/43,400), **nado dvermi** 1% (214/34,300).

7.2.1.4. **odo**

SSJ'59 says that **odo** is used in the collocation **odo mňa** *from me*, and more seldom alongside **od** before some other words, e.g. **dvere** *door*, **deň** *day*. The

collocation **odo dverí** *from the door* occurs among the examples. KSSJ'97 and PSP'00 say that **odo** is used in the collocation **odo mňa** and before forms of the words **dvere** *door*, **deň** *day*, **dvor** *courtyard*, **dno** *bottom*.

In the corpus investigated, a few other collocations also occurred with a significant frequency of vocalisation. For comparison, the corresponding data retrieved from SNK are presented in the following table.

odo dna	114	13%	od dna	737
odo dňa	3,824	92%	od dňa	311
odo dnes	47	8%	od dnes	561
odo dneška	223	59%	od dneška	153
odo dverí	302	27%	od dverí	813
odo dvier	76	62%	od dvier	46
odo dvora	6		od dvora	68
odo mňa	11,663	99%	od mňa	87

Table 70. SNK data as of 14 August 2012 regarding collocations with **odo**.

The data from the different sources confirm each other to a very high degree, but the relative rate of **odo dverí** *from the door* and **odo dvora** *from the courtyard* is significantly lower in SNK. **Odo mňa** *from me* is practically obligatory and **odo dneška** *from today* is almost equally as frequent as **od dneška**, whereas **odo dvora** is rare, which is also confirmed by Google data: **odo dvora** 1% (279/21,900). The two different forms of **dvere** *door* are interesting, since the rate of vocalisation of the less common, more exceptional form, is higher than with the more regular form. This difference is also confirmed by Google data, even though the relative rate of both forms with the vocalised preposition is significantly lower: **odo dverí** 11% (3,660/30,300), **odo dvier** 21% (401/1,530).

7.2.1.5. podo

SSJ'59 says that **podo** is only used in the collocations **podo mňa** *under me* (direction), **podo mnou** *under me* (position), **podo dvermi** *under the door* (position), whereas KSSJ'97, PSP'00 and PSP'05 say that **podo** is used in the collocations **podo mňa**, **podo mnou**, **podo dvermi**, **podo dvere** *under the door* (direction). For comparison, the corresponding data retrieved from SNK are presented in the following table.

podo dvere	1		pod dvere	59
podo dverami	0		pod dverami	75
podo dvermi	2		pod dvermi	4
podo mňa	23		pod mňa	1
podo mnou	603	100%	pod mnou	1

Table 71. SNK data as of 14 August 2012 regarding collocations with **podo**.

It seems evident that the only collocations in frequent use with the vocalised form of this preposition are those involving the first person singular personal pronoun. All collocations except **podo mnou** are quite infrequent in SNK and for comparison the following data were retrieved with Google: **podo mňa** 41% (2,900/4,230), **podo dvere** 1% (294/46,600), **podo dverami** 0% (0/13,100), **podo dvermi** 1% (37/3,020). It should especially be noted that according to the Google data, there is a severe drop in the rate of vocalisation in the collocational type involving the accusative case of the pronoun denoting direction towards a place *under or beneath me*. All the other collocational types involving the vocalised form of the preposition must be characterised as rare. The rate of **podo dvermi** is slightly higher in SNK, but the absolute number of occurrences is too small to be statistically significant.

7.2.1.6. **predo**

SSJ'59 says that **predo** is used in the collocations **predo mnou** *in front of me* (position), **predo mňa** *in front of me* (direction), more seldom instead of **pred** before some other words, e.g. **dvere** *door*, **deň** *day*. It is also said that **predo dňom** *before daybreak* is synonymous with **pred svitaním** *before dawn* and it is recommended to write it as one word: **predodňom**. To write **predo dňom** as two separate words is said to be old fashioned. KSSJ'97, PSP'00 and PSP'05 say that **predo** is used in the collocations **predo mňa**, **predo mnou**, **predo dvere** *in front of the door* (direction), **predo dvermi** *in front of the door* (position), **predo dňom**. They do not contain any entry for **predodňom**. *Slovník slovenských synonymov* (SSS'95) mentions that **predo dvermi** is synonymous with **v malej časovej vzdialenosti** *in a small distance of time*, i.e. that it constitutes a fixed phrase with a temporal meaning of *just around the corner*. For comparison, the corresponding data retrieved from SNK are presented in the following table.

predo dňom	129	20%	pred dňom	512
predodňom	0			
predo dvere	149	27%	pred dvere	397
predo dverami	80	5%	pred dverami	1,682
predo dvermi	961	93%	pred dvermi	72
predo dvercami	0		pred dvercami	0
predo dverce	1		pred dverce	0
predo mňa	573	98%	pred mňa	8
predo mnou	5,951	99%	pred mnou	50

Table 72. SNK data as of 14 August 2012 regarding collocations with **predo**.

The collocations with the personal pronoun seem to be practically obligatory. The collocational types **predo dňom** and **predo dvere** are less frequent than

their counterparts with the non-vocalised form of the preposition. For collocations with the instrumental case **dver(a)mi**, there seems to be something very close to a complementary distribution, so that **predo dvermi** and **pred dverami** are the two preferred combinations, whereas **pred dvermi** and **predo dverami** are quite rare. Also, the instances with the irregular shorter form **dvermi** are much fewer than instances with the regular form **dverami**, and a quick investigation of some of the data clearly shows that all four collocational types involving **pred(o) dver(a)mi** may be used in both the literal spatial meaning *in front of the door, right outside the door, on the doorstep*, as well as in the metaphorical temporal meaning *just around the corner, soon*. Hence there is no obvious semantic difference conditioning the use of the individual collocational types.

7.2.1.7. spodo

This vocalised form is not mentioned in any of the dictionaries investigated. For comparison, analogical collocations with the first person singular personal pronoun have been retrieved from SNK.

spodo mňa	3	spod mňa	7
-----------	---	----------	---

Table 73. SNK data as of 14 August 2012 regarding collocations with **spodo**.

The relative rate of the vocalised form is very low in SNK, but a Google search for **spodo mňa** (118/30) *from underneath me* results in a rate of 80%, hence it seems that this preposition is treated analogous to other prepositions in collocations containing this pronoun. The somewhat lower rate may be caused by the fact that these collocations have a very low absolute frequency, hence they do not become as well encoded in the speakers' minds as other more frequent collocations with this pronoun. The rate still seems high enough, however, for the codification to prescribe the use of the vocalised form.

7.2.1.8. spredo

SSJ'59 says that **spredo** is used in the collocation **spredo mňa** and alongside **spred** before some other words, e.g. **dvere** *door*, **deň** *day*, whereas KSSJ'97, PSP'00 and PSP'05 say that **spredo** is used in the collocations **spredo mňa** *from in front of me*, **spredo dverí** *from in front of the door*, **spredo dňa** *from before the day*.

For comparison, the corresponding data retrieved from SNK are presented in the following table.

spredo dňa	0	spred dňa	1
spredo dverí	9	spred dverí	50
spredo dvier	2	spred dvier	1

spredo mňa	8	spred mňa	0
------------	---	-----------	---

Table 74. SNK data as of 14 August 2012 regarding collocations with **spredo**.

There are very few occurrences even in SNK, and the data were therefore supplemented by a search with Google which gave the following data: **spredo dňa** 0% (1/1,610), **spredo dverí** 8% (373/4,400), **spredo dvier** (5/45), **spredo mňa** 100% (671/1). Vocalisation is clearly obligatory before the first person singular personal pronoun, whereas **spredo dverí** is infrequent, and **spredo dňa** as well as **spredo dvier** are rare.

7.2.1.9. znado

SSJ'59 says that **znad** is synonymous with **sponad** and that the form **znado** is rarely used. Examples are given which contain **znado dvier** and **znad mňa**. KSSJ'97, PSP'00 and PSP'05 do not contain the entry **znad**.

For comparison, the corresponding data retrieved from SNK are presented in the following table.

znado dverí	0	znad dverí	0
znado dvier	2	znad dvier	0
znado mňa	0	znad mňa	1

Table 75. SNK data as of 14 August 2012 regarding collocations with **znado**.

As there is so little data in SNK, a Google search was made and it gave the following results: **znado dverí** (1/0), **znado dvier** (4/0), **znado mňa** (0/9). These very few data confirm the standpoints in the dictionaries, and the extremely low frequency of this preposition is most likely due to a preference for the synonymous form **sponad** *from above*.

7.2.1.10. navzdory

All the dictionaries investigated list both **navzdor** and **navzdory** without any comment on the difference between them. In SNK 1,496 instances of **navzdory** vs. 52 of **navzdor** gives a vocalisation rate of 97%. Google's 158,000 instances of **navzdory** vs. 10,500 of **navzdor** gives a total vocalisation rate of 94%. It therefore seems that **navzdory** is the standard form in Slovak, and that the non-vocalised form is relatively rare.

7.2.1.11. skrze

All the dictionaries investigated list both **skrz** *through* and **skrze** without any comments on the difference. SSJ'59 and KSSJ'97 say that this preposition is somewhat old-fashioned. In SNK there are 5,818 instances of **skrze** and 1,884 instances of **skrz**, which gives a rate of 76%. It seems, however, that the majority of the non-vocalised instances occur in the collocation **skrz naskrz** *all*

through, from end to end, hence just as with **navzdory**, the vocalised form **skrze** seems to be the standard form in modern Slovak, whereas the non-vocalised form is far less frequent, with the exception of the phrase **skrz naskrz**.

7.2.1.12. **ku**

SSJ'59 says that **ku** is used before words beginning with ⟨k⟩ or ⟨g⟩ or a consonant cluster containing ⟨k⟩ or ⟨g⟩, more rarely also before other consonant clusters, and also in the collocation **ku mne**. Besides examples of the usage before ⟨k⟩ and ⟨g⟩, the collocations **ku vkladu** and **ku cti** occur in the examples. KSSJ'97 says that **ku**, pronounced [ku] or [gu], is used before ⟨k⟩ and ⟨g⟩, before consonant clusters containing ⟨k⟩ or ⟨g⟩, and sometimes also before other consonants, as well as in the collocation **ku mne** where the preposition is always pronounced [ku]. Besides examples of its usage before ⟨k⟩ and ⟨g⟩, the collocations **ku skrini** and **štyri ku dvom** also occur in the data. PSP'00 and PSP'05 have no comments on the usage of the form **ku**.

Strangely, nothing is said about the rate with which the vocalised form of the preposition is used in the different contexts mentioned. The findings from the corpus will therefore be compared with data from SNK in order to try and make some generalisations.

Already, based on the previous corpus investigation and the statements in the dictionaries, there can be no doubt that it is highly appropriate to characterise the vocalised form **ku** as obligatory before words beginning with the phonemes /k/ and /g/. That it is a simplification to talk about graphemes, as the dictionaries seem to do, is however easily demonstrated by rather large numbers of instances such as, e.g. **ku Clintonovi** [gu'klintonovi] *to Clinton* and **k gintonicu** [gdʒin'toniku] *to the gin tonic*. A very small number of true counterexamples were however found in the corpus, but almost all of them involved words of foreign origin, and more often they begin with ⟨g⟩ than with ⟨k⟩.

With regard to the next major context, i.e. before consonant clusters containing /k/ or /g/, but not in the first position, the collocational types mentioned in the literature as well as the relevant data exhibiting any significant frequency in the corpus make it difficult to detect any more specific patterns. It is however evident that the number of actually occurring consonant clusters with any significant frequency is quite restricted, and that clusters with /g/ are extremely rare.

More general searches in SNK based on the type of initial consonant cluster also confirm the earlier pattern identified in the corpus, namely that in these contexts, the vocalised form of the preposition is used just about equally as often as the non-vocalised form of the preposition. It should however be pointed out that individual collocational types with a certain lexeme may have a somewhat higher or lower relative frequency than the average for the initial consonant cluster in question.

ku sk*	4,114	51%	k sk*	3,973
ku šk*	814	56%	k šk*	644
ku vk*	52		k vk*	42
ku vzk*	39		k vzk*	20
ku zk*	2		k zk*	3

Table 76. SNK data as of 14 August 2012 regarding collocations with **ku** before consonant clusters containing /k/ not in the initial position.

Moving on to other instances of the vocalised form **ku** before consonant clusters not containing /k/ or /g/, data regarding the following collocational types have been collected from SNK because they are mentioned in the literature or because exhibit a significant frequency of vocalisation in the corpus.

ku cti	917	98%	k cti	16
ku dnešnému	68	6%	k dnešnému	1,074
ku dnešnej	2	1%	k dnešnej	221
ku dnešným	2	1%	k dnešným	153
ku dnu	777	96%	k dnu	29
ku dnám	1		k dnám	1
ku dňu	1,870	97%	k dňu	52
ku dňom	6		k dňom	13
ku dverám	1,138	16%	k dverám	5,772
ku hre	102	13%	k hre	690
ku hrám	13		k hrám	64
ku hranici	32	2%	k hranici	1,349
ku hraniciam	11	2%	k hraniciam	487
ku hvieзде	8		k hvieзде	72
ku hvieздам	219	41%	k hvieздам	311
ku mne	19,497	100%	k mne	3
ku Slovensku	74	8%	k Slovensku	870
ku snímke	6		k snímke	60
ku snímkam	0		k snímkam	55
ku svätej	23	11%	k svätej	190
ku svätému	32	6%	k svätému	478
ku svätým	10	8%	k svätým	113
ku svätosti	86	21%	k svätosti	331
ku svojej	644	11%	k svojej	4,973
ku svojmu	542	6%	k svojmu	8,664
ku svojim	634	10%	k svojim	5,606
ku špičke	59	11%	k špičke	498

ku špičkám	5		k špičkám	36
ku šťastiu	347	36%	k šťastiu	624
ku štátu	15	3%	k štátu	511
ku štátom	14	7%	k štátom	184
ku štyrom	80	18%	k štyrom	363
ku trom	108	8%	k trom	1,293
ku včerajšiemu	62	18%	k včerajšiemu	284
ku včerajšej	4		k včerajšej	30
ku včerajším	5		k včerajším	66
ku vchodu	511	60%	k vchodu	339
ku vchodom	9		k vchodom	8
ku všetkému	1,180	77%	k všetkému	362
ku všetkej	12		k všetkej	7
ku všetkým	2,222	71%	k všetkým	917
ku vzdelaniu	157	23%	k vzdelaniu	529
ku vzniku	506	22%	k vzniku	1,779
ku vzťahu	85	35%	k vzťahu	158
ku vzťahom	30	29%	k vzťahom	73
ku zmene	237	5%	k zmene	5,020
ku zmenám	96	5%	k zmenám	1,889
ku zneužitiu	4	2%	k zneužitiu	242
ku zrážke	32	8%	k zrážke	371
ku zrážkam	10	7%	k zrážkam	141
ku zvieraťu	7		k zvieraťu	86
ku zvieratám	68	13%	k zvieratám	462

Table 77. SNK data as of 7 August 2012 regarding collocations with **ku** before consonant clusters not containing /k/.

From the above table it becomes clear that the vocalised form **ku** is used in a very wide range of instances, but with a rather low rate of vocalisation. It seems that there might be a general tendency to vocalise before consonant clusters at a rate of 5–25%. Besides that general tendency, a few collocational types exhibit vocalisation at a much higher rate. Vocalisation in the collocations **ku mne** *towards me*, **ku cti** *to the honour*, **ku dnu** *to the bottom*, **ku dňu** *to the day* seems to be practically obligatory. Vocalisation is also highly frequent before forms of the pronoun **všetok** *all*. That it is conditioned by the cooccurrence with this specific lexeme, and not by the initial consonant cluster, is demonstrated by the data in the following table, where the rate of vocalisation before other lexemes beginning with the same cluster is just a normal 11%.

ku vř*	3,557	60%	k vř*	2,404
ku vřetk*	3,424	73%	k vřetk*	1,292
ku vř*	133	11%	k vř*	1,112
but not vřetk*			but not vřetk*	

Table 78. SNK data as of 7 August 2012 regarding collocations with **ku** before words beginning with ⟨vř⟩.

Among the remaining data, the collocations **ku hviezdám** *towards the stars*, **ku vchodu** *towards the entrance* are equally as frequent as the corresponding collocations with the non-vocalised form of the preposition.

The remaining contexts exhibit only slight variations, which may be explained as individual choices in phonetically complex contexts due to the rather free variation in the vocalisation of this preposition.

7.2.1.13. so

SSJ'59 states that **so** is pronounced [zɔ], with the exception of the collocation **so mnou** *with me*, where it is pronounced [sɔ]. There is also an example containing the collocation **so vřetkým** *with all*. KSSJ'97 says only that **so** is pronounced [zɔ] except in the collocation **so mnou**. No further comments on the use of **so** are given in any of the dictionaries.

The previous investigation of the corpus data led, however, to some rather clearly defined contexts which trigger vocalisation on phonological grounds, as well as a number of other instances of a seemingly more lexical character. The only somewhat problematic context is before ⟨v⟩ followed by a sibilant or an affricate consonant. In order to confirm the earlier findings, the corresponding data were retrieved from SNK.

so vř*	292	55%	s vř*	235
so vs*	2,443	99%	s vs*	37
so vř*	21,776	99%	s vř*	255
so vz*	6,996	98%	s vz*	127
so vř*	51		s vř*	34

Table 79. SNK data as of 14 August 2012 regarding collocations with **so** before /vř/ and /vř/.

It becomes evident that vocalisation is practically obligatory before ⟨vs⟩, ⟨vř⟩, ⟨vz⟩, but is close to being equally as frequent as the non-vocalised form before ⟨vř⟩ and ⟨vř⟩.

Other, non-phonological vocalisations with a significant frequency in the corpus have, for comparison, been retrieved from SNK and are presented in the following table.

so <i>čtou</i>	816	98%	s <i>čtou</i>	18
so <i>ctihodným</i>	7		s <i>ctihodným</i>	5
so <i>džbáňkom</i>	0		s <i>džbáňkom</i>	3
so <i>lž*</i>	33		s <i>lžou</i>	7
so <i>mnou</i>	29,565	100%	s <i>mnou</i>	7
so <i>mzdou</i>	69		s <i>mzdou</i>	5
so <i>mzdami</i>	81		s <i>mzdami</i>	5
so <i>mzdov*</i>	50		s <i>mzdov*</i>	21
so <i>psom</i>	2,236	100%	s <i>psom</i>	1
so <i>psami</i>	1,146	99%	s <i>psami</i>	8
so <i>psík*</i>	308	95%	s <i>psík*</i>	18
so <i>psi*</i>	132	92%	s <i>psi*</i>	12
so <i>psou</i>	35		s <i>psou</i>	4
so <i>psovod*</i>	82		s <i>psovod*</i>	8
so <i>psoriázou</i>	56		s <i>psoriázou</i>	18
so <i>pstruh*</i>	20		s <i>pstruh*</i>	0
so <i>psychológ*</i>	209	51%	s <i>psychológ*</i>	203
so <i>pšenicou</i>	47	38%	s <i>pšenicou</i>	77
so <i>pštros*</i>	12		s <i>pštros*</i>	12

Table 80. SNK data as of 14 August 2012 regarding collocations with **so**.

From the above data it seems rather clear that vocalisation is practically obligatory in the collocational types **so mnou** *with me*, **so psom** *with the dog*, **so psami** *with the dogs*, and close to obligatory in the collocational types **so čtou** *with honour*, **so mzdou** *with the salary*, **so mzdami** *with salaries*, as well as in collocations containing derivations from the root {ps} *dog*. Furthermore, a very high rate of vocalisation is also exhibited by collocations containing derivations from the roots {mzd} *salary* and {lž} *lie*. The vocalised form of the preposition is almost equally as frequent as the non-vocalised form in collocations containing derivations beginning with **cti-** *honour*. Vocalisation in the collocational type **so džbáňkom** *with the jug* is rare.

The data concerning collocations with words beginning with ⟨ps⟩ and ⟨pš⟩ are more difficult to interpret. Here the variation in the rate of vocalisation differs between the individual collocational types. The total numbers of vocalisations before these two consonant clusters, except for instances with the root {ps} *dog*, are shown in the following table.

so <i>ps* other than the lexeme pes</i>	668	35%	s <i>ps* other than the lexeme pes</i>	1,232
so <i>pš*</i>	86	40%	s <i>pš*</i>	129

Table 81. SNK data as of 14 August 2012 regarding collocations with **so** before ⟨ps⟩ and ⟨pš⟩.

7.2.1.14. vo

None of the dictionaries gives any information on the usage of the vocalised form of the preposition **vo**. Apart from the obviously obligatory vocalisations before words beginning with /v/ or /f/, SŠJ'59 contains the following collocations among the examples: **vo dne v noci** *day and night, all around the clock*, **necítiť sa vo svojej koži** *not feel quite all right*, **vo štvrtok** *on Thursday*, **zložiť papier vo dvoje** *to double fold a piece of paper*, **vo dva konce skrčená** *completely bent over*; somewhat old fashioned: **vo svet široký** *into the wide world*; and marked as archaic: **volať vo tri vrhy** *call three times*, **pozdraviť vo troje** *greet threefoldly*, **vo meno božie** *in the name of god*, **žiť si vo svet** *to lead a comfortable life*. KSSJ'97 repeats a few examples from SŠJ'59, whereas SSS'95 adds the example **vo dve poly** *double folded*.

The corpus data showed a very high degree of vocalisation (94–97%) before consonant clusters containing a sibilant followed by ⟨v⟩, as well as the cluster ⟨štv⟩. For all other collocational types containing a ⟨v⟩ or ⟨f⟩ as the second or third element in the following word, mentioned by the dictionaries or exhibiting a significant frequency in the corpus, the corresponding data have been retrieved from SNK. These data are presented in the following table.

vo dva	46		v dva	49
vo dve	18		v dve	24
vo dvoch	2,008	10%	v dvoch	17,531
vo dverách	3,667	94%	v dverách	244
vo dvoje	252	97%	v dvoje	7
vo dvojici	1,363	83%	v dvojici	281
vo dvojiciach	270	54%	v dvojiciach	228
vo dvojhre	2,741	75%	v dvojhre	918
vo dvojhrách	144	39%	v dvojhrách	228
vo dvorane	55	38%	v dvorane	88
vo dvoranách	1		v dvoranách	4
vo dvore	1,203	97%	v dvore	32
vo dvoroch	96	82%	v dvoroch	21
vo hvezdárni	48		v hvezdárni	15
vo hviezde	18		v hviezde	1
vo hviezdach	1,026	98%	v hviezdach	20
vo hviezdň*	64	33%	v hviezdň*	132
vo chválach	5		v chválach	14
vo chvíli	7,498	99%	v chvíli	41
vo chvíľach	1,666	98%	v chvíľach	35
vo chvíľke	50		v chvíľke	18
vo chvíľkach	29		v chvíľkach	4
vo chvíľkov*	24		v chvíľkov*	25

vo Chvojnici	14		v Chvojnici	6
vo chvoste	5		v chvoste	26
vo Lvove	173	55%	v Lvove	143
vo Rwande	14	2%	v Rwande	553
vo rwandsk*	5		v rwandsk*	41
vo sfére	1,635	85%	v sfére	281
vo sférach	79		v sférach	16
vo skvel*	158	16%	v skvel*	803

Table 82. SNK data as of 14 August 2012 regarding collocations with **vo** before selected onsets containing /v/ or /f/.

Because of the large variation in the rate of vocalisation before identical consonant clusters, it is improbable that these vocalisations are phonologically triggered. Therefore such collocational types, with a high absolute and relative frequency, will probably have to be regarded as part of the Slovak lexicon. Collocations with low frequency can be explained as a smaller general tendency to vocalise before lexemes with a non-initial /v/ or /f/ in the initial consonant cluster.

Continuing to vocalisation before clusters without /v/ or /f/, the following collocations are mentioned in the dictionaries or exhibit a significant frequency in the corpus. For comparison, data from SNK are presented in the following table.

vo dne	2,728	— ⁸⁰	v dne	30
vo meno	44	39%	v meno	70
vo mňa	421	100%	v mňa	2
vo mne	13,152	100%	v mne	11
vo mzde	127	63%	v mzde	76
vo mzdách	1	1%	v mzdách	106
vo mzdov*	1	0%	v mzdov*	213
vo sne	3,430	83%	v sne	682
vo snách	81	45%	v snách	100
vo snoch	10	2%	v snoch	553
vo štyri	2		v štyri	35
vo tri	57	54%	v tri	48
vo troje	23		v troje	1

Table 83. SNK data as of 14 August 2012 regarding frequent collocations with **vo** before onsets not containing /v/ or /f/.

⁸⁰ All the instances with the non-vocalised form of the preposition contain the noun **dno** *bottom*, whereas the majority of the collocations with the vocalised form of the preposition contain the noun **deň** *day*.

In this group of collocations there is no obvious reason for any phonologically triggered vocalisations, hence these collocations must be interpreted as part of the lexicon. **Vo mňa** *into me* and **vo mne** *in(side) me* follow the general pattern of vocalisation before all forms of this pronoun. The vocalisation in the collocation **vo dne** *in the day(s)*, based on the lexeme **deň** *day*, seems to be obligatory as it serves to express a lexical distinction from the lexeme **dno** *bottom*, which occurs in the non-vocalised collocation **v dne** *in the bottom*. The vocalisation in **vo mzde** *in the salary* is only a little more frequent than the corresponding non-vocalised collocation. The collocational type **vo sne** *in the dream* is very frequent, and the corresponding plural **vo snách** *in dreams*, involving an irregular case ending, is almost equally as frequent as its non-vocalised counterpart, whereas **vo snoch**, with the regular case ending, is rare, maybe analogically based on the lexicalised irregular **vo snách**. In a few instances the vocalised form of the preposition only occurs in longer set phrases, namely **vo meno božie** *in the name of God*, **vo tri vrhy** *three times*, **vo tri smery** or **vo tri strany** *in three directions*, or in slight variations of these phrases, e.g. **vo meno Otca, Syna i Ducha Svätého** *in the name of the Father, the Son and the Holy Spirit*. The vocalisation in the collocation **vo troje** *threefold(edly), in three pieces, three persons together* seems to be practically obligatory.

Even if the dictionaries do not make any comment on the use of the vocalised form before consonant clusters containing /v/ or /f/ in the second or third position, this type of context is mentioned with regard to the other syllabic prepositions. On the background of the previous findings in the corpus examined, all collocations in which the preposition **v(o)** is followed by a consonant cluster containing an initial fricative, followed by a second or third element which is /v/ or /f/⁸¹, were retrieved from SNK.

vo sv*	190,328	95%	v sv*	9,523
vo zv*	8,862	81%	v zv*	2,045
vo šv*	6,735	92.5%	v šv*	548
vo žv*	0	0%	v žv*	2
vo chv*	9,340	95%	v chv*	450
vo hv*	1,196	85%	v hv*	208
vo sf*	1,773	83%	v sf*	357
vo stv*	93	24%	v stv*	298
vo skv*	165	15%	v skv*	962
vo schv*	45	8%	v schv*	492
vo zdv*	28	22%	v zdv*	98

⁸¹ The total set of such initial clusters, according to Sawicka (1974: 96ff.).

vo štv*	60,123	90%	v štv*	6,665
vo škv*	6	17%	v škv*	30

Table 84. SNK data as of 14 August 2012 regarding collocations with **vo** before consonant clusters containing an initial fricative and /v/ or /f/ as second or third element.

It turns out that vocalisation is highly frequent before a sibilant or a velar fricative followed by /v/ or /f/. Before clusters containing an intervening plosive the vocalisation rate is about 20%, with the exception of a much higher relative frequency before the cluster (štv). This, however, is best explained as a lexicalisation involving certain word roots, above all the lexemes **štvrtok** *Thursday* and **štvrtý** *fourth* and derivations thereof.

7.2.1.15. zo

SSJ'59 says that the preposition **zo** is used in the collocations **zo mňa** *from me*, **zo dňa** *from the day*, **z(o) dverí** *from the door*, and before words beginning with the consonants ⟨z⟩, ⟨s⟩, sometimes also ⟨ž⟩, ⟨š⟩, or with a consonant cluster containing ⟨z⟩, ⟨s⟩. The following collocations were further mentioned in the article: **zo dňa na deň** *from a day to another*, **zo dňa 15. augusta** *from the 15th of August*, **nebude zo psa slanina** *nothing good will come out of it*, **strhnúť zo mzdy** *to deduct from the wages*, **zo všetkých strán** *from all directions*, **zo všetkých síl** *with all one's energy*, **vymotať sa zo všetkého** *to worm out of everything*, **najhorší zo všetkých** *the worst of all*. The dictionary also contains a separate entry **zo** for the approximative particle used before numerals, stating that the form **z** is occasionally also encountered in this function. KSSJ'97 also reports that the vocalised form is used before words beginning with ⟨s⟩, ⟨z⟩, ⟨š⟩, ⟨ž⟩, and sometimes also before consonant clusters containing these consonants, as well as in the collocations **zo mňa** *from me*, **zo dňa** *from the day*, **z(o) dverí** *from the door*. There is also a separate entry for the lexeme **zo** functioning as a particle in the approximative meaning *some, about*. In the article a few collocations are repeated from SSJ'59. SSS'95 adds the collocations **zo všetkých končín** *from all parts*, **zo všetkých kútov sveta** *from every corner of the world*. PSP'00 and PSP'05 do not make any comments.

The corpus data investigated showed a vocalisation rate of over 99% before all the four sibilants, i.e. confirming what is said in KSSJ'97 rather than in SSJ'59. The same high rate of vocalisation was also found before /v/ followed by a sibilant, but not before other types of consonant clusters containing a sibilant in the second or third position, hence the situation is more complicated than is reported by the dictionaries.

zo vs*	1,206	99%	z vs*	17
zo vš*	31,277	100%	z vš*	103

zo vz*	7,432	100%	z vz*	46
zo vž*	37		z vž*	6

Table 85. SNK data as of 16 August 2012 regarding collocations with **zo** before selected clusters containing a sibilant in the second position.

Hence, it follows from this that the vocalised form **zo** is only used before consonant clusters consisting of /v/ followed by a sibilant, and that the vocalised form is obligatory in this context, with a possible exception for /vž/, which however also very frequently triggers the vocalised form in SNK.

Therefore, except for the contexts which obligatorily trigger vocalisation, data regarding all collocational types mentioned by the dictionaries or exhibiting a significant frequency in the corpus have been extracted from SNK. These data are presented in the following table.

zo Bzenca	0		z Bzenca	2
zo Bzenova	2		z Bzenova	10
zo lži	83		z lži	9
zo lží	19		z lží	6
zo lživ*	2		z lživ*	2
zo mzdy	1,013	100%	z mzdy	7
zo mzdov*	67		z mzdov*	13
zo ps* <i>dog</i>	382	90%	z ps*	42
as noun	320	98%	as noun	9
as adj.	62		as adj.	33
zo pseudo*	10		z pseudo*	15
zo psycholog*	75	30%	z psychologá	169
zo pšeni*	58	29%	z pšeni*	141
zo pštros*			z pštros*	
as noun	11		as noun	2
as adj.	19		as adj.	38
zo xenofóbie	7		z xenofóbie	8
zo Xibalby	0		z Xibalby	0

Table 86. SNK data as of 16 August 2012 regarding collocations with **zo** before selected clusters containing a sibilant.

It seems that vocalisation is practically obligatory before the nouns **mzda** *salary*, **pes** *dog* and **psík** *little dog*, but not before derived adjectives, where vocalisation is merely the more frequent option. Vocalisation is also notably frequent before the nouns **lož** *lie* and **pštros** *ostrich*. It seems, however, to be a general tendency that the rate of vocalisation is lower before derived adjectives than before the corresponding noun.

Because of the considerable variation in vocalisation rates before different

clusters in the above table, it does not seem likely that these vocalisations may be triggered phonologically. On the other hand, the number of collocational types is quite high, and it seems problematic to include them all in the lexicon of standard Slovak. Instead, it seems that a smaller number of close to obligatory collocations should be considered part of the lexicon, and a principle dealing with this type of initial consonant clusters makes individual variation possible in these contexts.

We will now proceed to vocalisations before consonant clusters not containing a sibilant. The following collocational types were mentioned in the dictionaries or exhibited a significant frequency in the corpus. For comparison, data from SNK are presented in the following table.

zo dna	238	22%	z dna	841
zo dňa	11,501	96%	z dňa	448
zo dverí	83	8%	z dverí	940
zo dvier	5	4%	z dvier	128
zo mňa	7,625	100%	z mňa	1
zo včely	0		z včely	3
zo včiel	7		z včiel	9
zo včera	287	91%	z včera	29
zo včerajška	220	97%	z včerajška	7
zo včerajšieho	400	79%	z včerajšieho	106
zo včerajšej	150	78%	z včerajšej	43
zo včerajších	121	86%	z včerajších	20

Table 87. SNK data as of 16 August 2012 regarding collocations with **zo** before selected clusters not containing a sibilant.

The total number of collocations with **zo vč*** in SNK is 1292, vs. **z vč*** 313, which gives a rate of 80%. Deducting the number of collocations in the above table leaves us with 107 **zo vč*** and 96 **z vč*** in other collocational types, with a rate of 53%. It seems, therefore, that both forms of the preposition are equally frequent before the consonant cluster /vč/, but that the lexicalisation **zo včerajška** is practically obligatory, whereas vocalisation in other collocations containing the root {včer} is just frequent. Further, **zo dňa** *from the day* is obligatory, whereas **zo dna** *from the bottom* and **zo dverí** *from the door* are infrequent and **zo dvier** *from the door* is rare.

7.2.2. Language Correctness

Language correctness handbooks are not as frequent in Slovakia as in Poland, but a few titles have appeared in recent years. The following four have short notes on the issue investigated.

Mistrík (1971: 98) says that asyllabic prepositions receive an extra /o/ when

the beginning of the following word is the same as or similar to the consonant constituting the preposition. This happens in order to facilitate pronunciation.

Pisárčiková et al. (1988: 170f.) say that prepositions which end in a consonant vocalise before the personal pronoun *ja I*. The asyllabic prepositions also vocalise before other words, but only: 1) when the following word begins with the same or almost the same consonant (differing only with respect to voice and/or alveolarity⁸²); 2) when the following word begins with a consonant cluster, where an identical or similar consonant is present in the second or a later position. For the preposition *bez without*, it is specifically mentioned that it vocalises in the collocation *bezo mňa without me*, but not under any other circumstances. The collocational types *bezo sporu without dispute*, *bezo zmeny without change*, *bezo zvyšku completely, without remainder* are not considered to belong to standard literary Slovak.

Mistrík, Škvareninová & Hegerová (1997: 65) say that the asyllabic prepositions vocalise when their shape coincides with the beginning of the following word. Vocalisation also occurs when the following word begins with a consonant cluster where one of the consonants is identical to the preposition. Furthermore, the prepositions *bez without*, *cez through*, *nad above*, *od from*, *pod under* vocalise in the collocations *bezo mňa without me*, *nado dvere above the door*, *odo dvier from the door*, *podo dvere under the door*, *vo dne in the day*, *zo dňa from the day*, *vo sne in the sleep, in the dream*. It is also spelled out that vocalisation does not occur before pronouns other than the pronoun *ja I*.

Horák (1999: 17f.) says that the vocalised form is pronounced [ku] only in the collocational type *ku mne*. In all other instances, i.e. before words that begin with ⟨k⟩ or ⟨g⟩, but also ⟨ch⟩ and ⟨h⟩, as well as with a consonant cluster containing one of the consonants ⟨k⟩, ⟨g⟩, ⟨ch⟩, ⟨h⟩, it is pronounced as [gu]: *ku komore towards the chamber*, *ku gajdám towards the bagpipes*, *ku chorému towards the sick*, *ku horárni towards the gamekeeper's cottage*, *ku skúške towards the exam*, *ku glazúre to the frosting*, *ku chrobákovi to the beetle*, *ku zhorenisku towards the place of conflagration*. In written texts, however, we sometimes see the non-vocalised form *k: k chate towards the cottage*, *k skupinke towards the group*, *k chladiacim towards the cooling*. In these instances, one should write *ku*.

All the handbooks with the exception of Mistrík (1971) talk about vocalisation before clusters containing identical or similar sounds in the second (or third) position. As they do not say anything about the rate of vocalisation in the different contexts, but simply state that the prepositions vocalise in the contexts mentioned, one has the impression that this is always the case in all the contexts. The data investigated, however, show clearly that this is not the case, i.e. that there are considerable differences in the rate of vocalisation in the

⁸² I.e. voiced or voiceless, pre-alveolar or post-alveolar.

different contexts mentioned. Furthermore, some of the aforementioned lexicalised collocations vocalise only rarely. It should be quite important in these kinds of works to emphasise the most frequent lexicalised collocations and possibly to provide information about their frequency. On the other hand, the vocalisations of **bez**, which are rejected by Pisárčiková et al. (1988), have proven to be quite frequent in the corpus data.

Horák's statements are interesting as he seems to argue for the obligatory use of the vocalised form of the preposition **k(u)** before all instances of ⟨k⟩, ⟨g⟩, ⟨ch⟩, ⟨h⟩, as well as before consonant clusters containing the corresponding sounds. According to the corpus data, this is quite far from the actual state of affairs, as actual usage has proven to be considerably more restricted.

7.2.3. Grammars

Reference grammars and textbooks on Slovak grammar are considerably more concerned with the issue under investigation than are the dictionaries and language correctness handbooks. Almost every modern grammar has some notes on the usage of the vocalised forms of the prepositions.

Vážný (1948: 118f.), in his Slovak grammar for Czech students, in addition to a number of collocational types that we have already discussed, such as **vo dne** *in the day*, **zo sna** *from the sleep*, **ku cti** *to the honour*, **ku mnohým** *towards many*, **ku vzniku** *to(wards) the rise*, **ku všetkému** *towards all*, points out the two collocations **ku podivu** *astonishingly enough* (lit. *to the astonishment*) and **ku Prahe** *towards Prague*. In a note he also says that the preposition [sic!] **so** very often expresses an approximative measure, e.g. **bolo nás so desať** *we were about ten*, **vypil so dva litre** *he drank some two litres*.

In SNK the rate of vocalisation of **ku podivu** (18/6) exhibits a very low absolute frequency and it can therefore be suspected to be a bohemism. **Ku Prahe** must also be a bohemism, as there is not a single such instance in SNK (0/59).

Letz (1950: 50) mentions **vo dne** *in the day*, **so psom** *with the dog*, **ku mne** *towards me*, **zo sna** *from the sleep*, **zo mňa** *from me*, and further says that the preposition is always vocalised if its consonant is identical or very similar to the initial consonant of the following word. If there is no such similarity, vocalisation still occurs before consonant clusters, especially before clusters which are difficult to pronounce. In other words, this phenomenon is not regular and one can observe extensive variation in the language, e.g. **k(u) svojim** *towards one's own*, **k(u) bráne** *towards the gate*. In a footnote it is added that the vocalisation is appropriate when it coincides with the strengthening of an even-numbered yer, e.g. **so psom** < *съ рьсьтъ with the dog, **vo dne** < *въ дъне *in the day*, but this appropriate vocalisation is often neglected, e.g. **v dnešnom svete** *in today's world*, **v psej nálade** *in a bad mood* (lit. *in a dog's mood*). On the other hand, vocalisation is not appropriate when it is not a reflex of a

yer, but an epenthetic vowel, e.g. **vo vode** *in the water*, **so slamou** *with straw*. Certainly, such vocalised prepositions support the melodiousness of the language, but it would still be unsuitable to apply them by force.

It is good to see that Letz stresses the fact that variation can be observed before consonant clusters. It is however alarming that he labels certain vocalisations as appropriate and others as inappropriate on historical grounds. The final statement actually gives the impression that he does not wish to recommend the usage of the vocalised form even in positions where such forms are used practically obligatorily in modern standard Slovak.

Pauliny, Štolc & Ružička (1955a: 280f.) say that the asyllabic prepositions and the prepositions **od**, **nad**, **pod**, **pred**, **bez**, **cez** vocalise when the following word begins with the same consonant or with a consonant cluster which makes the pronunciation difficult. Apart from some obvious instances, the following collocations are mentioned: **odo dňa** *from the day*, **ku škole** *towards school*, and sometimes also **ku otcovi** *towards the father*, **ku materi** *towards the mother*, **ku stromu** *towards the tree*. Vocalisation always occurs before forms of the pronoun **ja I**.

Pauliny, Ružička & Štolc (1955b: 66) say that the asyllabic prepositions **s**, **z**, **k**, **v** and the syllabic prepositions **od**, **nad**, **pod**, **pred**, **bez**, **cez** vocalise when the following word begins with the same consonant or with a group of consonants before which it would be difficult to pronounce the final consonant of the prepositions. Vocalised prepositions always occur before forms of the pronoun **ja**.

These two statements are quite unfortunate, as talking about vocalisation of syllabic and asyllabic prepositions on equal terms, the scope becomes far too broad. Also, the instances **ku otcovi** and **ku materi** are highly surprising as no evidence has been found in the data that vocalisation of **k** before words beginning with a vowel or a single consonant other than /k/ and /g/ are applied with any regularity in standard Slovak.

Zauner (1958: 62f.) says that the asyllabic prepositions always vocalise before words beginning with the same consonant as the one in the preposition, or with the corresponding voiced or voiceless consonant; **s** and **z** also vocalise before words beginning with **š** or **ž**. The asyllabic prepositions often also vocalise when the following word begins with a consonant cluster containing a consonant which is identical to or similar to the one in the preposition. Among the examples are **vo dvore** *in the courtyard*, **ku sklárni** *towards the greenhouse*, **ku školníkovi** *to the school caretaker*, **v(o) dverách** *in the door*, **z(o) pšenice** *from wheat*. In other instances vocalisation occurs only seldom. Asyllabic and syllabic prepositions vocalise before the pronoun **ja I** as well as in the collocations **vo dne** *in the day*, **vo sne** *in the sleep, in the dream*, **zo dňa** *from the day/date*, **od(o) dňa** *from the day*, **nad dverami/nado dvermi** *above the door*, **pred dverami/predo dvermi** *in front of the door, outside the door*.

This statement is quite well formulated, but still not precise. Also,

according to the corpus data, the collocations **vo dvore** and **vo dverách** should not be presented as instances of phonologically triggered vocalisation, but rather as lexicalisations, as the general rate of vocalisation before ⟨dv⟩ is quite low.

Uhlár (1960: 126) says that the asyllabic prepositions vocalise in order to improve pronunciation. Further, **od, nad, pod, pred, bez, cez** always vocalise before forms of the pronoun **ja**, as well as in some other set expressions, e.g. **predo dvermi, predo dvere, odo dňa**. It is pointed out that only the non-vocalised form is used in e.g. **bez sporu, bez slov, bez zmyslu**.

Unfortunately nothing is said about the rate of vocalisation with regard to the set expressions, and Uhlár's standpoint concerning the use of the non-vocalised form **bez** is obviously prescriptive, as corpus data have shown that some collocational types containing the form **bezo** are quite frequent.

Ružička et al. (1966: 631ff.) say that the prepositions constitute a phonetic unit together with the following word. Monosyllabic prepositions have a common main stress together with the following word. Therefore some prepositions change their form depending on the following word. 1) All primary prepositions which end in a consonant vocalise before the pronoun **ja**. 2) Vocalised forms of the prepositions **nad, od, pod, pred** occur in some fixed collocational types before forms of the words **dvere, deň, dvor, dno**, e.g. **predo dvermi** *in front of the door, outside the door*, **predo dňom** *before the day(break)*, **nado dvere** *above the door*, **nado dňom** *before daybreak; (God reigned) over the day(light)*, **odo dvora** *from the courtyard*, **odo dňa** *from the day*, besides the non-vocalised **pred dverami, nad dvere**, etc. In the standard language the vocalised forms of the prepositions **bez, cez** are not used before words beginning with a sibilant, therefore e.g. **bezo zvyšku** *without remainder*, **bezo zmeny** *without change* are not standard forms. 3) The asyllabic prepositions vocalise if the following word begins with the same consonant as the one in the preposition, or with the corresponding voiced/voiceless consonant, and for the prepositions **s, z** also with the postalveolar sibilants **š, ž**. In the written language, however, in contrast with the spoken language, when the preposition **z** is followed by a word beginning with **ž**, the preposition is sometimes left unvocalised. 4) The asyllabic prepositions are also vocalised before words beginning with a consonant cluster containing, in its second or third position, a consonant that is identical or similar (differing only with regard to voice and type of alveolarity) to the preposition. Sometimes, however, such collocations occur without vocalisation. Also, the preposition **v** vocalises less frequently than **s, z**. Some collocational types only occur with the non-vocalised form of the preposition, e.g. **tvárou v tvár** *face to face*, **v tvojom** *in your*, **v tvrdzi** *in the fortress*. 5) The preposition **k** vocalises with the highest frequency of all prepositions. In an artistic style, authors from central and eastern Slovakia use **ku** almost without any limitations. 6) Vocalised prepositions also occur in some instances without regard to the following

consonants, e.g. **zo dňa na deň** *from one day to another*, **so dňom** *with the day*, **z(o) dvora** *from the yard*, **z(o) dvier** *from the door*, **vo dne** *in the day*, **v(o) sne** *in the sleep/dream*, **vo snách** *in the dreams*, **v(o) tme** *in the dark*.

This presentation is precise in most respects, but it lacks some detail. The statement about the usage of the vocalised form **ku** in regionally coloured language is very enlightening. One of the examples is however not very relevant, as with a rate in SNK of 1% (2/138), the collocation **so dňom** is clearly rare.

PSP'67 (p. 24f.) mainly copies the presentation in Zauner (1958). No interesting new details are presented with the exception of the statement that the vocalised form of the preposition is always used in **vo sne**. SNK only exhibits a vocalisation rate of 83%.

Pauliny, Štolc & Ružička (1968: 306) say that the asyllabic prepositions vocalise when the following word begins with an identical or articulatorily related consonant, or with a consonant cluster, before which it would be difficult to pronounce the asyllabic preposition. The forms **ku otcovi** *to the father*, **ku matere** *to the mother* and **ku sestre** *to the sister* are mentioned as sometimes used. Asyllabic prepositions, as well as the syllabic prepositions **od, nad, pod, pred, bez, cez**, vocalise before forms of the pronoun **ja I**. To facilitate pronunciation, vocalisation also occurs in the collocation **odo dňa** and in the adverbials **nadovšetko, predovšetkým**, but not when the prepositions are used freely, e.g. **nad všetko**.

Pauliny (1971: 130f.) says that the asyllabic prepositions vocalise before words beginning with the same consonant as the consonant in the preposition, as well as before words beginning with a similar consonant. This is followed by examples like **zo vzduchu** *in the air*, **ku skale** *towards the cliff*, **vo dvore** *in the yard*. With forms of the pronoun **ja I**, every preposition (except some compound prepositions) must end with a vowel. Alongside **pred dverami, nad dverami, od dňa**, the collocations **predo dvermi, nado dvermi, odo dňa** are also correct.

These formulations are very short. Unfortunately, contexts with a similar consonant in the second position of a cluster are not mentioned, only included in the examples. Further, the mention of just a few fixed collocations gives the impression that this list might be exhaustive.

Bartoš (1972: 209) says that prepositions are vocalised in order to facilitate pronunciation.

Bělič et al. (1980: 133), in their textbook on the Slovak language for Czech students, say that the asyllabic prepositions, as well as the syllabic ones, vocalise in certain positions in order to achieve a better pronunciation.

Oravec & Laca (1980: 180f.), in their handbook on orthography, say that in order to facilitate the pronunciation, the asyllabic prepositions vocalise before words beginning with the same consonant as the one in the prepositions, or its voiced or voiceless counterpart. The prepositions **s, z** also vocalise before **š, ž**.

The asyllabic prepositions further vocalise before a consonant cluster where the second or third consonant is identical to the consonant in the preposition or differs only with regard to voice and/or type of alveolarity. However, this is often not manifested in the written language with regard to the preposition **v**, e.g. **v(o) svetle** *in the light*, **v(o) svojom** *in one's own*, **v dvojstupe** *in two rows*, **v dvojjici** *in the couple*, **v dvanástich** *in the twelve*, **v tvojom** *in your*, **v tvare** *in the shape*. The preposition **v** should not be vocalised in contexts not complying with the aforementioned principles, e.g. **v škole** *in the school*, **v stole** *in the table*, **v stene** *in the wall*, **v meste** *in the town*. On the other hand the preposition **k** is frequently used in more contexts than predicted by the principles mentioned, e.g. **k(u) hmle** *towards the fog*, **k(u) stromu** *towards the tree*, and even before single consonants, e.g. **k(u) sestre** *to the sister*. The asyllabic as well as the syllabic prepositions vocalise before forms of the personal pronoun **ja I**. The following collocations also occur: **pred dverami/predo dvermi** *in front of the door*, **nad dverami/nado dvermi** *above the door*, **od(o) dňa** *from the day*, **v(o) sne** *in the sleep, dream*, **vo snách/v snoch** *in the dreams*, **so cfou** *with honour*. The syllabic prepositions do not vocalise even if the following word begins with the same consonant as the final consonant of the preposition, e.g. **bez zvyšku** *without remainder*, **bez zmeny** *without change*, **bez sporu** *without dispute*.

Again, the scope of the principles given is not precisely delimited and the rate of vocalisation in different contexts is not made clear enough. The collocation **ku hmle** is interesting. It is very infrequent in SNK (3/1). The rate of vocalisation retrieved with Google (9/384) indicates that **ku hmle** is very rare in everyday modern Slovak.

Pauliny (1981: 201f.) has almost the same wording as Pauliny (1971), but adds that the asyllabic prepositions also vocalise if a consonant which is identical or closely related to the consonant in the preposition occurs in the initial consonant cluster of the following word.

Mistrík (1983: 98), in his Slovak grammar published in English, says that the asyllabic prepositions vocalise “when the following word begins with the same or a similar consonant or when the non-vocalised preposition would be difficult to pronounce [...]. The vocalised forms are always used with the forms of the personal pronoun *ja* [...]. Vocalised forms are also to be found with the prepositions *od, nad, pod, pred, bez, cez: odo mňa, nado mňa,*” etc.

Mistrík (1984: 140) says that the asyllabic prepositions vocalise when there would otherwise arise consonant clusters that would be difficult to pronounce. With forms of the pronoun **ja I**, prepositions end with a vowel: **odo mňa, bezo mňa, pred o mňa**.

Findra et al. (1986: 134f.) say that prepositions constitute one unit with the following word and share a single common stress with that word. Therefore the form of some prepositions changes depending on the form of the following word. 1) All primary prepositions ending with a consonant, i.e. **bez, cez, nad, od, pod, pred, s, v, z, k**, vocalise before the pronoun **ja I**. 2) The vocalised

forms of **nad, od, pod, pred** occur in some set collocational types with the words **dvere, deň, dvor, dno**. The examples are the same as in Ružička et al. (1966). Vocalised forms of **bez, cez** are not used in standard Slovak other than after the pronoun **ja**. Collocations such as **bezo zmeny** *without change*, **bezo zvyšku** *without remainder* are sub-standard. 3) The asyllabic prepositions vocalise before words beginning with the same consonant as the one in the preposition or with its voiced/voiceless counterpart. The prepositions **s, z** also vocalise before words beginning with **š, ž**. 4) The asyllabic prepositions also vocalise before words beginning with a consonant cluster containing an identical or corresponding voiced/voiceless consonant in the second or third position.

Again the scope of the last principle seems to be far too broad and the rate of the individual types of vocalisation is not problematised.

Oravec, Bajzíkova & Furdík (1988: 182f.) say that the form of the prepositions alternates depending on the phonetic characteristics of the initial sounds in the following word in order to facilitate the pronunciation. All the prepositions **pred, nad, od, pod, bez, cez, s, z, v, k** vocalise before the personal pronoun **ja**. The asyllabic prepositions vocalise: 1) when the following word begins with the same consonant as the one in the preposition or with a consonant only differing with respect to voice or type of alveolarity; 2) when the following word begins with a consonant cluster where the second or third consonant is identical or only differs with regard to voice or type of alveolarity. Only exceptionally is the non-vocalised form of the prepositions permitted, namely before the stops **t, d**, e.g. **v tvare** *in the form*, **v dvanástej** *in the twelfth*. In individual instances the syllabic prepositions also vocalise before consonant clusters beginning with an identical consonant, e.g. **predo dvermi** *in front of the door*, **predo dňom** *before the day*, **odo dňa** *from the day*. There are also other individual reminiscences of an older norm for the vocalisation of prepositions, e.g. **vo dne** *in the day*, **vo sne** *in the sleep/dream*, **v(o) snách** *in the dreams*, **v(o) tme** *in the dark*.

It is interesting to see that the maximal interpretation of the vocalisation in the context before clusters containing an identical or similar consonant in the second or third position is made more explicit through the subsequent delimitation, made by means of the overruling principle which permits the non-vocalised preposition before /t/ and /d/. It is not said explicitly, but this exception may refer only to the preposition **v(o)**, as the type of context described is practically impossible with regard to the other asyllabic prepositions. The statement that the type of initial consonant in the following word would influence the vocalisation of syllabic prepositions as well is also quite problematic.

Papierz (1994: 65), in her Slovak grammar for Polish learners, says that the asyllabic prepositions vocalise before a following word beginning with an identical or similar consonant to the consonant in the preposition. The forms

so and **ku** are pronounced [zo] and [gu], with the exceptions of **so mnou**, **ku mne**.

Mistrík (1994: 79) says that the asyllabic prepositions are vocalised when a consonant cluster would otherwise arise which would be difficult to pronounce. With forms of the pronoun **ja I**, prepositions always end in a vowel.

Bujalka & Dubníček (1998: 147) say that vocalisation of prepositions facilitates their pronunciation. The asyllabic prepositions regularly vocalise in certain contexts, whereas the syllabic prepositions only do so in exceptional cases. The asyllabic prepositions vocalise if the following word begins with an identical consonant or with a similar consonant only differing with respect to voice or type of alveolarity or if an identical or similar consonant appears in the second or third positions of the initial consonant cluster of the following word. All prepositions vocalise before forms of the pronoun **ja I** and the nouns **dvere door** and **deň day**. Vocalisation is incorrect in collocations like **bezo zvyšku without remainder**.

Although basically correct, once more the principles are not narrowed down to the exact contexts and the rate of vocalisation in the different contexts is not problematised.

Navrátil (2001: 53f.) says that prepositions vocalise in order to facilitate pronunciation, because when words meet [sic!], Slovak does not allow (completely or partially) alike consonants to follow after each other. The asyllabic prepositions vocalise almost regularly, whereas **bez**, **cez**, **nad**, **pod**, **pred**, **od** vocalise less often. All the prepositions mentioned vocalise before forms of the pronoun **ja I**. The asyllabic prepositions vocalise if the following word begins: a) with an identical consonant to the one in the preposition; b) with an articulatorily similar consonant with regard to voice and/or type of alveolarity; c) with a consonant cluster where the second or third consonant is identical or similar except for voice and/or type of alveolarity.

PSP'05 (p. 136f.) says prepositions constitute one rhythmical unit together with the following word, and therefore, depending on the initial sound of the following word, the preposition assimilates or vocalises. Vocalisation occurs with the asyllabic prepositions **s**, **z**, **v**, **k** in order to facilitate their pronunciation. 1) The asyllabic prepositions are vocalised when the following word begins with the same consonant as the one in the preposition, or its counterpart with regard to voice. 2) The prepositions **s** and **z** vocalise if the following word begins with **š** or **ž**. 3) Vocalisation also occurs if the following word begins with a consonant cluster where one consonant is identical or similar (with regard to voice and/or type of alveolarity) to the consonant in the preposition. 4) The asyllabic prepositions are vocalised before personal pronouns: **so mnou**, **ku mne**, **vo mne**, **zo mňa**. 5) Before the personal pronoun **ja I**, the monosyllabic prepositions **od**, **pod**, **pred**, **nad**, **bez**, **cez** also vocalise. 6) The syllabic prepositions do not vocalise if the following word begins with

the same consonant as the final consonant of the preposition. However, the vocalised form of the preposition may be used alongside the non-vocalised form in the collocations **pred dverami/predo dvermi** *in front of the door*, **nad dverami/nado dvermi** *above the door*, **od(o) dňa** *from the day*, **v dni/vo dne** *in the day*, **v(o) snách** *in the dreams*, **v(o) sne** *in the sleep/dream*, **z(o) dňa** *from the day*, **bez(o) zmeny** *without change*, but only **so čou** *with honour*.

This presentation contains much the same information as most of the previous works. Point number 4 is, however, probably by accident, far too broad, as it speaks about pronouns in general. The most interesting point is the permissive example of possible vocalisation in the collocation **bezo zmeny**, which has been condemned as sub-standard by all other authors.

7.2.4. Linguistics

As with Polish, the interest in vowel-zero alternations in prepositions is quite modest. The corresponding alternations in prefixes get much more attention. There are actually only a couple of articles and some shorter passages in a couple of monographs on different aspects of the subject.

Ryba (1934) clearly points out that Havlík's rule cannot be applied to modern Slovak.

Letz (1937) speaks about the increasing euphony achieved through vocalisation of prepositions. He refers to Ryba's (1934) three proposed types of vocalisations: the historically expected vocalisations, instances of neglected historically expected vocalisations, and innovative vocalisations by analogy. He then poses the highly interesting question *Why is that so?* Why are the historically expected realisations in the minority? He says the cause is to be found in the fact that prepositions do not form a fixed unit with the following word, and refers to Haller's (1934) investigation of vocalisation of Czech prepositions in texts where tendencies, but no firm principles, can be traced. He expects similar results for Slovak, i.e. that Havlík's rule does cope with the present state and that firm principles will be difficult to define. With regard to the asyllabic prepositions he presents some historically correct vocalisations, e.g. **vo dva konce** *folded in half* (lit. *in two ends*), **so všetkými** *with all*, **vo dverách** *in the door*, **vo sne** *in the sleep/dream*, **vo mne** *in(side) me*, **vo hre** *in the play*, **so psami** *with the dogs*, **ku dverám** *towards the door*, **ku ľvovi** *towards the lion* (but better: **k ľavovi**), **ku hre** *to(wards) the play*, **ku cti** *to the honour*, **zo dverí** *from the door*, **zo všetkého** *from all*, **zo psa** *from the dog*, **zo⁸³ dňa na deň** *from one day to another*; as well as analogical vocalisations, e.g. **vo svete** *in the world*, **vo Viedni** *in Vienna*, **ku takým** *to(wards) such*, **ku prijatým** *to the accepted*, **ku jej bozku** *to her kiss*, **ku zázrakom** *to the miracles*, **ku salonu**

⁸³ According to the orthography of that time: **so dňa na deň**.

towards the salon, **ku Beránkovi** to *Beránok*, **ku banke** to(wards) the bank, **ku tomu** to that, **ku nemu** to him, but he does not explain the principles according to which analogy works. He also points out that **so sebou** with oneself is correct; **s sebou** should be avoided. With regard to the syllabic prepositions, vocalisation is normally neglected, and they are most significantly preserved before forms of the first person personal pronoun **odo mňa**, **bezo mňa**, etc., as well as, by analogy, **odo dávna**, **odo dvora**, **bezo zmien**. Similarly to Haller's findings, it can be stated for Slovak that vocalisation never occurs if the following word begins with a vowel, and the asyllabic prepositions are always vocalised if the following word begins with the same consonant as the one in the preposition. Apart from that, we find a "motley promiscuity" in the usage of vocalised and non-vocalised forms of prepositions. There is even variation in identical contexts, e.g. **v(o) svojej**, **k(u) nemu**, **v(o) dverách**, and "it would be erroneous to wish to search by force for rules with many exceptions for the vocalisation of prepositions as a whole, because it would not simplify the question in any way". Mainly because more vocalisations give a higher degree of euphony, Letz recommends **s(o) ženou** with the woman, **ku vnučke** to the grandchild, **vo dvore** in the yard, **vo štvrtok** on Thursday, **vo svete** in the world, **vo snách** in the dreams, **ododávna** since long ago, **nadovšetko** above all, but no arguments are given for this standpoint. The general recommendation is to vocalise for euphony where one has a choice between the two forms of the preposition, but there is still the possibility of using the non-vocalised form in styles where euphony is not so important. One should not use the vocalised form where it is against the general tendencies, e.g. before a consonant followed by /r/ or /l/.

It is evident that the historical background to the present situation still plays an important role in discussions of the issue, even though this author is well aware that the historical principles cannot serve as guidelines for vocalisation in modern Slovak. Instead, he favours vocalisation whenever it is possible. The problem then remains how to know whether it is possible or not. Unfortunately, Letz does not believe that any systematical pattern for vocalisation of prepositions can be found in modern Slovak.

Peciar (1947), in his article about asyllabic prepositions in Slovak, makes something of a step forward in the direction of answering the question of when vocalisation of syllabic prepositions occurs in standard Slovak. His reasoning is based on the regularities for vocalisation of Czech asyllabic prepositions presented by Haller (1934). Peciar links his explanation of the vocalisation of asyllabic prepositions to the synchronic principles for consonant clusters in Slovak in general, and claims that vocalisation takes place in such instances where lack of vocalisation would lead to a consonant cluster that may not occur initially, neither in simple nor in compound Slovak words.

This model of explanation is of course highly relevant and extremely interesting, but it would need to be accompanied by a precise investigation

and description of the principles defining the possible structures of word-initial syllable onsets in standard Slovak.

Sabol (1966) only investigates the two prepositions **k** and **v**. He states that they never vocalise before a following word beginning with a vowel, and that instances like **ku otcovi** *to the father* are rather dialectal or at least socially marked. Such vocalisations should be regarded as inappropriate as they go against the natural Slovak principle to avoid hiatus. He also states that **v** always vocalises before 'f' and 'v', whereas **k** always vocalises before 'k' and 'g'. In all other instances a certain variation can be observed. There are no significant differences in the usage of the vocalised and non-vocalised forms of the preposition in data collected from prose and poetry, respectively. There is, however, a significant difference in the rate of vocalisation between the two prepositions investigated. The distribution of the preposition **k(u)** is less dependent on the shape of the following word than is the distribution of **v(o)**. The forms of the preposition **k(u)** are more freely used and this preposition is more frequently vocalised than the preposition **v(o)**. These differences could be explained by the acceptability of initial consonant clusters in Slovak. Because of the prefix 'v-', Slovak is much more used to and accepts more complex onsets beginning with 'v-' than with 'k-', as there exists no prefix corresponding to the preposition **k**. In the data investigated, the vocalised preposition **vo**, except before 'f' and 'v', only occurs before consonant clusters, usually with 'v' as the second consonant. The traditional collocations **vo mne**, **vo dne**, **vo meno božie**, **vo sne** have to be interpreted as synchronic anomalies, and in all these instances a certain tendency can be observed towards using the non-vocalised form. The tendency for **k(u)** is the opposite. A tendency can be observed towards vocalising in all kinds of contexts, before clusters without 'k' or 'g', as well as before single consonants other than 'k' or 'g', or even before words beginning with a vowel. Sabol says that with regard to **k(u)**, standard handbooks ought to be more daring and more favourable towards the vocalised form **ku**. He finishes by saying that since the reason for vocalisation is to be sought in the set of allowed onset types in Slovak, one can expect to find similar results for the prepositions **s** and **z** as for the preposition **v**, because of the existence of the corresponding prefix(es)⁸⁴ 's-/z-'.

Oravec (1968: 20ff.) is the only linguistic monograph concerned with the whole spectrum of this issue in Slovak. He gives a concise survey of the problem in a book which is otherwise mainly concerned with the syntax and semantics of prepositions. Ten Slovak prepositions, i.e. **bez(o)**, **cez(o)**, **k(u)**,

⁸⁴ Historically and semantically, it can be argued that there are two prefixes. They have however become phonologically neutralised, and they are both realised by the same two allophones [s] and [z]. In modern standard Slovak orthography, they are both spelled phonetically as <s> and <z>, respectively.

nad(o), od(o), pod(o), pred(o), s(o), v(o) and **z(o)**, exhibit final vowel–zero alternation. Since the preposition and the following word constitute a common phonetic unit with one common primary stress, the preposition alternates depending on the phonetic character of the following word, either by assimilation or by vocalisation. He divides the prepositions into four asyllabic and six syllabic. The asyllabic prepositions vocalise before words beginning with the same consonant as the one in the preposition, or almost the same, differing only with regard to voice and/or type of alveolarity. Vocalisation also occurs if the relevant consonants are not at the very beginning of the following word, but in the second position of a consonant cluster. **S** and **z** may also vocalise if the second consonant is an affricate, e.g. **zo včelína** *from the beehouse*. **K** vocalises more frequently than the principles predict, e.g. **ku vchodu** *towards the entrance*, **ku cti** *to the honour*, **ku vzdeleniu** *to the education*, **ku stene** *towards the wall*, **ku chlebu** *to(wards) the bread*, **ku hrobu** *towards the grave*, **ku boku** *towards the side*, **ku pomoci** *to help*, **ku deťom** *to the children*. The preposition **v**, on the other hand, vocalises to a somewhat more restricted degree than predicted by the principles, e.g. **v dvojici** *in the couple*, **v tvári** *in the face*, **v kvete** *in the flower*, i.e. not before clusters with an initial ‘t’, ‘d’, ‘k’. In written language, it is often also not vocalised when the initial consonant is a sibilant, e.g. **v svojom** *in one’s own*, **v svetle** *in the light*, but in spoken language it is normally vocalised. Exceptional vocalisations also occur, e.g. in **zo cti** *of the honour*. Furthermore, vocalisation of all ten prepositions occurs before forms of the pronoun **ja** *I*, as well as sometimes in the prepositions **nad, pred, od, v, s, z** before the words **dvere, dvor, deň, tma** in the collocations **vo dverách** *in the door*, **vo dvore** *in the yard*, **vo dne** *in the day*, **v(o) tme** *in the dark*, **nado dvere** *above the door*, **predo dvermi** *in front of the door*, **odo dňa** *from the day*, **predo dňom** *before the day*, **z(o) dvora** *from the yard*, **zo dňa** *from the day*, **z(o) tmy** *from the dark*. The preposition **bez** only vocalises in the collocation **bezo mňa**. **Bezo sporu** *without dispute*, **bezo zvyšku** *without remainder*, **bezo zmyslu** *without sense* and **bezo cti** *without honour* are sub-standard.

Kučera (1982) begins by stating that it is generally considered “that vocalisation of prepositions in Czech as well as Slovak is an exclusively phonological phenomenon which has no semantic differentiating function. Evidently in Slovak there is no expression whatsoever in which the use of the vocalised form of the preposition versus the basic form could serve to distinguish between meanings.” (ibid.: 55) The discussion of data in the article is however restricted to the context of asyllabic prepositions before words beginning with a single consonant followed by a vowel. Kučera reviews the statements in the traditional handbooks. He agrees that the asyllabic prepositions always vocalise before words beginning with the same consonant as the consonant in the preposition, or with its voiced/voiceless counterpart. He states, however, that the vocalisation of **s** and **z** before words beginning with **š-** and **ž-** followed by a vowel has been imprecisely formulated. According to

his investigation of a smaller amount of corpus data, this type of vocalisation occurs at a rate of 97.5%, which is significantly more than in the same context in Czech.

Rubach (1993: 277ff.) gives an account of the vocalisation of the asyllabic prepositions within the framework of lexical phonology. He states that this process is governed by “rules which apply in domains larger than words and which are hence postlexical. [...] The surfacing of the vowel in the prepositions [...] is due to constraints on permissible onsets. These constraints prohibit sequences of identical or near-identical consonants. If such sequences arise as a result of syntactic concatenations, then the consonant of the prepositions remains extrasyllabic. This situation is remedied by inserting a vowel.” The problem with this statement is “that the distribution of [o] and [u] in the prepositions is not readily predictable. [...] To solve this problem we argue that the prepositions are in fact represented as sequences of consonants and yers at the melodic level: //vO, zO, gU//.” Then there is “a rule of vowel vocalisation that inserts an X slot over the yer, that is, over the floating melody segment.” He goes on to say that “postulating yers in the prepositions is necessary independently in order to account for the form of the preposition in lexicalised phrases [...], for instance *zo dňa na deň* ‘from day to day’, *ku cti* ‘in honour’. The surfacing of the vowel in the preposition is due to Yer Vocalization⁸⁵ since the following noun has a yer [...]. These phrases are idiomatic. They are listed in the lexicon and hence they are subject to Yer Vocalization, which is lexical.” In a footnote he also adds that vocalisation occurs in instances like **ku škole** *towards the school* because “the coronal fricatives *s, z,* and *š* are transparent with regard to the constraints on permissible sequences of labials and velars.” This means that “they check the adjacency of the particular nodes that are dependent on PLACE. In the cluster /g šk-/ velar stops are adjacent at the DORSAL node.”

Vojteková (2011: 51ff.) says that it is only possible to explain the reasons for vocalisation of prepositions against the background of linguistic diachrony. She gives a very thorough presentation of the historical background of the forms of the Slavic prepositions as well as the forms they have taken in the individual modern Slavic languages. However, when it comes to the principles governing the choice between the available forms in Slovak and Polish, she merely repeats the most basic facts mentioned in the majority of handbooks during the last half century. She even refers to the principle of vocalisation when the non-vocalised form of the preposition would be difficult to pronounce, and gives, among others, the example **ku stene** *towards the wall*. When it comes to the syllabic prepositions she mentions that they only vocalise in set phrases with the words **dvere** *door*, **dvor** *courtyard*, **deň** *day*, **dno** *bottom*, and that the non-vocalised form of the preposition can be used as well.

⁸⁵ This rule is basically the same as Havlík’s law of yer drop or strengthening.

7.3. Summary of Slovak Data

The findings based on the corpus data and the review of the previous literature on the subject will now be summarised in order to give an easily accessible survey of the vocalisations which occur in modern standard Slovak.

Fifteen prepositions exhibit two alternating forms, with or without an additional final vowel. These prepositions fall into three groups: the syllabic **bez(o)**, **cez(o)**, **od(o)**, **nad(o)**, **pod(o)**, **pred(o)**, **spod(o)**, **spred(o)**, **znad(o)**, and the asyllabic **k(u)**, **s(o)**, **v(o)**, **z(o)**, which exhibit systematic patterns of vocalisation, and the two prepositions **skrz(e)** and **navzdor(y)**, which have not proven to exhibit such systematic patterns. Instead they almost exclusively occur in the longer shape **skrze** and **navzdory**. The preposition **znad** is very infrequent, since the synonymous, non-vocalising **sponad** is normally used in the same meaning.

It should also be noticed that **s** and **z** share the same basic phonological shape, /z/, and that the basic phonological shape of the preposition **k** is /g/, hence when these prepositions vocalise they acquire the shapes ⟨so⟩ /zo/, ⟨zo⟩ /zo/, and ⟨ku⟩ /gu/, respectively.

There is also a separate lexeme **zo** constituting an approximative adverbial particle which practically always occurs in this longer form regardless of the shape of the following word.

Just as in Polish, the grounds for vocalisation fall into two main categories: phonological and lexical. Only the four asyllabic prepositions vocalise on phonological grounds, whereas vocalised forms of both asyllabic and syllabic prepositions occur on lexical grounds. An important difference in comparison to Polish is that semantic differences between collocations with the vocalised form of the preposition and the corresponding collocations with the non-vocalised form are much less common in Slovak than in Polish.

The behaviour of the preposition **k** differs from all the other vocalising prepositions. On the one hand, there is clear statistical evidence that this preposition vocalises on certain phonological and lexical grounds, but on the other there are also significant but low rates of vocalisation in almost any context, even before words beginning with a vowel. The reason for this is the situation in central and eastern colloquial Slovak, where the form **ku** has outnumbered the asyllabic form **k**, just as in standard Polish. Under the influence of these dialects, the two forms **k** and **ku** seem to occur almost in free variation, but also to be sensitive to the complexity of the initial syllable structure of the following word, hence there is a tendency to vocalise more often before words beginning with longer consonant clusters than before a simple consonant or a vowel. The influences of this free variation are clearly noticeable in the corpus data and have an evident impact on actual usage in standard Slovak. Nowadays, however, this usage is discouraged by most of the newer codificational works, saying that this usage is not part of the

standard language, whereas many older works were quite permissive of it. Because of these diverging tendencies, which may all lead to the same result, it is in some instances difficult to tell which principle is actually at work with regard to the vocalisation of **k**.

7.3.1. Phonological Vocalisations

Purely phonologically triggered vocalisations of the prepositions **k**, **s**, **v** and **z** occur whenever the right phonological context is met, i.e. without any other reference to the following lexeme than its phonological structure.

In modern standard Slovak such phonological principles trigger obligatory vocalisation of the four asyllabic prepositions before words beginning with the same consonant as the one constituting the preposition or its counterpart with regard to voice, i.e. the vocalised form **ku** is used before /k/ or /g/, **vo** before /v/ or /f/, **so** and **zo** before /s/ or /z/. The two prepositions **s** and **z** also vocalise obligatorily before the corresponding post-alveolars /š/ and /ž/.

Furthermore, the prepositions **s** and **z** vocalise obligatorily before the consonant clusters /vs/, /vš/, /vz/, and the preposition **v** vocalises obligatorily before the consonant cluster /sv/. Also, vocalisation of **v** before /sf/, /šv/, /zv/ as well as **z** before /vž/ occurs at a very high rate, whereas the vocalised form **so** before /vž/ is slightly less frequent than **zo**. Before the cluster /žv/ the preposition **v** was not frequent enough for the data to be significant.

		before				
		/vs/	/vz/	/vš/	/vž/	
so		100%/99%	99%/99%	99%/98%	75%/60%	
zo		100%/99%	100%/99%	100%/100%	—/86%	
		/sv/	/zv/	/šv/	/žv/	/sf/
vo		97%/95%	94%/81%	97%/92%	?	76%/83%

Table 88. Relative rate of vocalisation in the corpus (left) and SNK (right).

A corresponding general high frequency of vocalisation of the preposition **k** before the corresponding consonant clusters /sk/, /zk/, /šk/, /vk/, /vzk/ does not exist.

		before				
		/sk/	/zk/	/šk/	/vk/	/vzk/
ku		40%/51%	—/40%	61%/56%	80%/55%	100%/66% ⁸⁶

Table 89. Relative rate of vocalisation in the corpus (left) and SNK (right).

⁸⁶ The high relative frequency before this cluster of three consonants might just as well be due to a general tendency exhibited by the preposition **k** to vocalise more often before longer consonant clusters.

This might be due to a general tendency to avoid vocalisations of this preposition in contexts other than before /k/ and /g/, based on statements in the codificational literature that discourage an exaggerated usage of the vocalised form **ku** under the influence of central and eastern Slovak dialects. Individual variation in the vocalisation rate before each type of consonant clusters seems to be highly dependent on the absolute frequency of the collocational type, so that the higher the absolute frequency of the collocational type, the higher also the relative frequency.

7.3.2. Vocalisations by Analogy

Possibly as an effect of the very broad principles formulated in almost every handbook, but maybe also as an effect of the wide range of different Slovak dialectal systems and the relatively short time during which Slovak has been standardised, there is a noticeable borderland between purely phonological vocalisations and purely lexicalised vocalisations.

In this borderland we find a quite varied set of collocations with different rates of the vocalised and the non-vocalised forms of the syllabic prepositions being used interchangeably before words beginning with certain consonant clusters containing the consonant found in the preposition, or a closely related consonant only differing with regard to voice and/or type of alveolarity as the second or third element. The consonant clusters involved are the ones before which the codificational literature allows or sometimes even recommends vocalisation, with the exception of the clusters already discussed above as always triggering vocalisation.

Because there is no general high rate of vocalisation before these consonant clusters, the vocalisations cannot be regarded as just phonologically triggered. The rate of vocalisation is low before certain word forms and higher before other word forms with the same initial cluster. It seems, however, that for each such consonant cluster there is one, or sometimes a few collocational types which exhibit a very high rate of vocalisation, often close to obligatory. Such collocations are best interpreted as lexicalisations. Alongside these collocations there is a number of contexts involving the same roots as the lexicalised collocational types. Such collocational types exhibit a significantly lower, but still quite high rate of vocalisation. They may best be treated as lexical vocalisations by analogy with the collocational types exhibiting obligatory or close to obligatory vocalisation. Finally, there are still other, lexically unrelated word forms beginning with the same consonant cluster. Many of these exhibit a very low but significant rate of vocalisation which may best be treated as vocalisation by phonological analogy with the initial consonant cluster of the collocational types exhibiting obligatory or almost obligatory vocalisation. This type of low rate vocalisation is supported by the

codificational literature, which practically unanimously states that vocalisation is always possible if the initial consonant cluster following the preposition contains the same consonant as the one in the preposition, or a similar consonant differing only with respect to voice and/or the type of alveolarity. Hence, there are two types of analogical vocalisation: lexical analogy, based on identity of roots, and phonological analogy, based on identity of the initial consonant cluster with a small group of lexicalised collocations exhibiting obligatory or close to obligatory vocalisation.

A good example is the consonant cluster /ps/, before which **so** occurs in SNK with a rate of 78%, and **zo** with a rate of 36%. It turns out, however, that vocalisation is obligatory in the collocations **so psom** *with the dog*, **so psami** *with the dogs*, **zo psa** *from the dog*, **zo psov** *from the dogs*, **so pstruhom** *with the trout*, **so pstruhami** *with the trouts*. Other combinations with the prepositions **s** and **z** followed by these two roots are vocalised somewhat less frequently, at a rate of 80-90% for nouns such as **zo pstruha** *from the trout*, **zo pstruhov** *from the trouts*, **zo psíka** *from the little dog*, **zo psíkov** *from the little dogs*, and 65-80% for adjectives such as **zo psieho**, **zo psej**, **zo psích** *from dog's*. The rate of vocalisation in collocations with other roots beginning with /ps/ varies between 25% and 75%, e.g. **so psoriázou** (SNK 76%) *with psoriasis*, **s psychologom** (SNK 53%) *with the psychologist*, **zo psychológie** (SNK 31%) *from psychology*. Hence it also seems that the rate of vocalisation is influenced by the absolute frequency of the collocational type in question, so that infrequent collocations also exhibit a lower relative rate of vocalisation.

A special category is also constituted by geographical names. The usage of the local population evidently has a great impact, so that **vo Chvojnici** is very frequent (SNK 70%), whereas **vo Chvatimechu** is not (SNK 0%).

Also, in some contexts where vocalisation would be expected, and which are often discussed in the literature, e.g. the preposition **v** before /tv/, there is no lexicalised collocational type with the vocalised form **vo**, hence no vocalisation may occur by analogy, which is confirmed by a rate of 0% in SNK. This strengthens the hypothesis that in order for a certain consonant cluster to be able to optionally trigger vocalisation, there must be a lexicalised collocation with a very high rate of vocalisation which encourages lexical as well as phonological vocalisation by analogy.

7.3.3. Lexical Vocalisations

All other instances of vocalisation, apart from the above-mentioned phonologically triggered ones, must be encoded in the lexicon of modern standard Slovak.

The following list constitutes a summary of the information in the literature and the corpus data which has been described earlier in this study. For obvious reasons it can be but a qualified approximation, as the data from the

different sources are sometimes somewhat contradictory.

7.3.3.1. Formal Lexicalised Vocalisations

Formal lexicalisations refer to instances of non-phonological vocalisation occurring in collocations where the semantics of the collocation does not differ from the sum of the semantics of its constituent parts, and where no limitations can be observed with regard to the contexts in which the collocation may be used, i.e. the same constituents do not occur in a non-vocalising collocation with another meaning or in certain other contexts. Of course, the optional formal lexicalised vocalisations occur with the non-vocalised form as well, but this option refers to the very same meaning and context.

7.3.3.1.1. Obligatory

The vocalised form is considered obligatory in the following lexicalised collocational types, according to several sources:

- before the pronominal forms **mňa/mne/mnou** *me*: **bezo mňa, cezo mňa, nado mňa, nado mnou, odo mňa, podo mňa, podo mnou, predo mňa, predo mnou, spodo mňa, spredu mňa, ku mne, so mnou, vo mňa, vo mne, zo mňa**;⁸⁷
- **ku cti** *to the honour*;
- **ku dňu** *to the day*; **vo dne** *in the day* (vs. **v dne** *in the bottom*), **zo dňa** *from the day*;
- **ku dnu** *to the bottom*,

as well as in the following collocations, and also analogically at a lower rate in derivations of the same root:

- **so cfou** *with the honour*;
- **vo dverách** *in the door*;
- **vo dvore** *in the courtyard*;
- **vo hvieзде** *in the star*, **vo hviezdach** *in the stars*,
- **vo chvíli** *in the moment*, **vo chvíľach** *in the moments*,
- **zo mzdy** *from the salary*;
- **so psom** *with the dog*, **so psami** *with the dogs*, **zo psa** *from the dog*, **zo psov** *from the dogs*;
- **so pstruhom** *with the trout*, **so pstruhami** *with the trouts*;

⁸⁷ The usage is however somewhat inconsistent in the less frequent collocational types. The percentage of vocalisation in the corpus (left) and SNK (right) is as follows: **bezo mňa** 94/100, **cezo mňa** 88/95, **nado mňa** 89/99, **nado mnou** 96/100, **odo mňa** 99/99, **podo mňa** 100/96, **podo mnou** 89/100, **predo mňa** 100/98, **predo mnou** 98/99, **spodo mňa** --/30, **spredu mňa** --/100, **ku mne** 100/100, **so mnou** 100/100, **vo mňa** 100/100, **vo mne** 100/100, **zo mňa** 100/100.

- **vo štvrtok** *on Thursday*;
- **zo včerajška** *from yesterday's day*.

7.3.3.1.2. Prevailing

Optional, but more frequent than the corresponding collocational type with the non-vocalised form of the preposition, are:

- **odo dňa** *from the day*;
- **vo Evove** *in Lviv*;
- **vo mzde** *in the salary*;
- **vo sne** *in the sleep, in the dream*;
- **ku všetkému**⁸⁸ *to(wards) all, to(wards) the whole*, **ku všetkej** *to(wards) the whole*, **ku všetkým** *to(wards) all*,

as well as the following collocations, and also analogically at a lower rate in derivations of the same root:

- **zo cti** *of honour*;
- **vo dvojici** *in a pair*;
- **so lžou** *with a lie*, **so lžami** *with lies*, **zo lži** *from the lie*, **zo lží** *from lies*;
- **so mzdou** *with the salary*, **so mzdami** *with the salaries*;
- **zo pstruha** *from the trout*, **zo pstruhov** *from the trouts*;
- **so pštrosom** *with the ostrich*, **so pštrosmi** *with the ostriches*, **zo pštrosa** *from the ostrich*, **zo pštrosov** *from the ostriches*;
- **ku schodom** *towards the stairs*;
- **vo štvrtfinále** *in the quarterfinal*;
- **vo štvrtom/štvrtrej/štvrtých** *in the fourth*;
- **zo včera** *from yesterday*;
- **ku vchodu** *towards the entrance*;
- **ku vzkrieseniu** *to the resurrection*.

7.3.3.1.3. Equally Distributed

Optional and equally as frequent as the corresponding collocation with the non-vocalised form of the preposition are:

- **odo dneška** *from today's day*,

as well as the following collocations, and also analogically at a lower rate in derivations of the same root:

- **ku hviezdám** *towards the stars*⁸⁹;

⁸⁸ The rate of vocalisation is somewhat higher before the masculine/neutre form **všetkému** than before the corresponding feminine **všetkej** and the plural **všetkým**. The reason seems to be a more frequent use of the vocalised form of the preposition before an independently used neuter form of this pronoun than before form followed by a head noun.

⁸⁹ The vocalised form seems to be somewhat more common in the sense *to a distant place*

- **vo štvrti** *in the district/quarter*;
- **zo včely** *from the bee*, **zo včiel** *from the bees*.

7.3.3.1.4. Infrequent

Optional and less frequent than the corresponding collocational type with the non-vocalised form of the preposition are:

- **ku dňom** *to the days*;
- **ku dverám** *to(wards) the door*, **predo dvere** (direction) *in front of the door*, **predo dvermi**⁹⁰ (position) *in front of the door, just around the corner*,

as well as the following collocations, and also analogically at a lower rate in derivations of the same root:

- **zo Bzenca** *from Bzenec*, **zo Bzenova** *from Bzenov*;
- **ku chate** *to(wards) the cottage*;
- **so pšenickou** *with wheat*, **zo pšenice** *from wheat*;
- **vo skvelom/skvelej/skvelých** *in the fantastic*;
- **vo stvorení** *in the creation*;
- **vo škvrnách** *in the stains*;
- **vo zdvihnutom/zdvihutej/zdvihnutých** *in the lifted*;
- **vo zdvorilosti** *in politeness*.

7.3.3.1.5. Rare

Optional and very infrequent are:

- **bezo cti** *without honour*;
- **predo dňom** *before the day*;
- **odo dnes** *from today*;
- **bezo dna** *without bottom*, **odo dna** *from the bottom*, **zo dna** *from the bottom*;
- **so dňom** *with the day*;
- **nado dvere** (direction) *above the door*, **nado dvermi**, **nado dverami** (position) *above the door*, **odo dverí**, **odo dvier** *from the door*, **podo dvere** (direction) *under the door*, **podo dvermi** (position) *under the door*, **pred dverami** (position) *in front of the door, just around the corner*, **spredo dvier**, **spredo dverí** *from in front of the door*, **zo dvier**, **zo dverí** *from the door*, **znado dvier**, **znado dverí** *from above the door*;
- **odo dvora** *from the courtyard*, **zo dvora** *out of the courtyard*;
- **vo mzdách** *in the salaries*;
- **bezo sna** *without sleep*;
- **vo snách**, **vo snoch** *in the dreams*;
- **vo tme** *in the dark*, **zo tmy** *from the dark*.

and somewhat less common in the sense *towards famous people*.

⁹⁰ SNK: **pred dverami** (60%), **predo dvermi** (34%), **pred dvermi** (3%), **predo dverami** (3%).

7.3.3.2. Contextual Lexicalised Vocalisations

In this category, most vocalised collocations are clearly semantically motivated, but not all. Some of them have a meaning that does not differ noticeably from the meaning of their constituent parts, but instead the collocation with the vocalised form is only used in certain restricted contexts. Contextual restrictions are of course always present with regard to semantic lexicalisations.

It is sometimes also the case that a contextual lexicalisation contains some morphological irregularities which are not present in the corresponding collocation with the non-vocalised form of the preposition, when used in the literal meaning of the collocation's constituent parts.

7.3.3.2.1. Obligatory

The following collocational types may be considered practically obligatory:

- **nado dňom** *before daybreak*;
- **vo meno božie** *in the name of God*;
- **vo dvoje** *folded in half; two persons together*, **vo troje** *threefold, folded three times; three persons together*, **vo štvoro** *folded four times; four persons together*;
- **nadovšetko, nado všetko** *above all, higher than anything* (not in the spatial meaning);
- **vo dva konce** *folded in half*;
- **vo dva vrhy** *twice*, **vo tri vrhy** *three times*⁹¹.

The obligatoriness of these collocations is either based on their very high rate of vocalisation or on implicit recommendations in the literature. However, they are not normally explicitly referred to as obligatory, but rather are reported to have a special status as only/always used in certain contexts.

7.3.3.2.2. Prevailing

Optional, but prevailing, are:

- **vo meno otca i syna i ducha svätého** *in the name of the Father, the Son, and the Holy Spirit*;
- **ku podivu** *astonishingly enough* (lit. *to the astonishment*);
- **bezo sporu** *without dispute*;
- **vo štyri sveta strany/kraje** *in all directions* (lit. *towards the four cardinal points*);
- **bezo zbytku** *without remainder*;
- **bezo zvyšku** *without remainder*.

⁹¹ Archaic, according to SSJ'59.

7.3.3.2.3. Infrequent

Optional and infrequent is:

- **ku pomoci** *to assistance, to help*;
- **bezo slov** *without a word* (lit. *words*);
- **ku šťastiu** *luckily* (lit. *to the luck*);
- **nado všetky** *beyond all* (in the non-spatial meaning, followed by a noun in the plural, e.g. **nado všetky pochybnosti** (SNK: 14%) *beyond all doubts*);
- **bezo zmeny** *without change*, **bezo zmien** *without changes*;
- **bezo zmyslu** *without any sense..*

8. The Czech Data

The number of alternating prepositions in Czech is somewhat larger than in Polish or Slovak. As in Polish, the alternating vowel in the majority of prepositions is /e/. As has been previously pointed out, Czech prepositions occur in their vocalised form to a significantly higher degree than prepositions in Slovak or Polish. The previous findings about the vocalisation of the Polish and Slovak prepositions will hopefully serve as a point of departure for the investigation of the Czech prepositions. Some parallel patterns could be expected to emerge, but at the same time, specific Czech patterns are also expected in order to explain the much higher frequency of vocalised prepositions in Czech.

As with the Polish and Slovak data, all instances of the vocalised form of the prepositions are extracted from the corpus, together with the following context word, and the possibility of using the non-vocalised form in the same contexts is then investigated. Based on these data, some preliminary hypotheses are formulated about the usage of the different forms of the Czech prepositions.

Next, the available literature on the subject is presented. The findings based on the corpus data are compared to the data presented in the literature. Similarities and differences between the corpus data and the literature are discussed, and finally, the results are summarised.

8.1. The Corpus Data

The seventeen Czech prepositions exhibiting vowel—zero alternation are listed in the table below. The number of occurrences in the corpus of the vocalised and non-vocalised form of each preposition is given, as well as the relative frequency of the vocalised form of each preposition.

Preposition	Vocalised		Non-vocalised
zpod(e)	0	0.0%	282
zpřed(e)	0	0.0%	1
pod(e)	124	0.3%	37,168
přes(e)	120	0.4%	33,335
před(e)	952	0.9%	99,877
nad(e)	630	1.5%	42,136
od(e)	2,337	1.6%	145,907
poblíž(e)	38	1.9%	2,013
bez(e)	1,816	3.6%	48,721
k(e)	38,061	13.1%	250,575
k(u)	2,421	0.8%	

v(e)	257,338	16.9%	1,264,585
z(e)	93,660	17.9%	428,587
s(e) ⁹²	97,532	18.3%	436,214
skrz(e)	926	74.1%	323
dík(y)	11,860	94.1%	737
krom(ě)	17,652	95.8%	776
navzdor(y)	2,727	97.0%	83
Total	528,194	15.9%	2,791,320

Table 90. Vocalisation in Czech prepositions.

The prepositions have been ordered according to their relative rate of vocalisation in the corpus. Once again, they clearly fall into three different groups: syllabic prepositions with a low frequency of the vocalised form (below 4%), asyllabic prepositions with a medium relative frequency of between 13% and 19%, and syllabic prepositions with a high frequency of the vocalised form (above 74%). The absolute frequency of the majority of the prepositions of the last type is low, however. Prepositions of this type were not present in the Polish data, but some of these prepositions did occur in the Slovak data, hence they do not constitute part of the common stock of Slavic prepositions. Finally, some of them vocalise by means of a vowel other than /e/, which is otherwise applied in the majority of Czech prepositions.

Just as with the Polish and Slovak data, the investigation will begin with the syllabic prepositions and then continue to the asyllabic ones.

8.1.1. Syllabic Prepositions

As in Polish and Slovak, it soon becomes clear that the syllabic prepositions are only found in collocations with a relatively small range of word forms each. The word forms encountered are presented below.

8.1.1.1. bez(e)

When it comes to the Czech preposition **bez(e)**, the number of collocational types is much higher than in Slovak or Polish. The 1,816 occurrences of the vocalised form are distributed over 128 different collocational types, whereof two were uninterpretable. 1,808 of the occurrences involve the vocalised form of the preposition followed by a consonant cluster. In the majority of cases there is a sibilant in the initial position, and only in a few cases in non-initial positions. A few instances also involve affricates containing a sibilant element,

⁹² This figure is based on a sample of 1000 instances of *se*, in which 87 were instances of the vocalised form of the preposition *s*. The figure therefore represents 8.7% of the 1,121,058 instances of *se*.

whereas 119 instances involve contexts without any sibilant element. There are also seven instances of a single initial consonant, of which three involve the sibilant /s/.

These brief data clearly indicate that Czech exhibits a richer variation than Slovak and Polish. It seems that individual choices, especially in fiction and poetry, may play an important role, based on the fact that such a high degree of single occurrences are encountered that they cannot all be explained as writing mistakes.

Therefore, a somewhat clearer picture emerges when we limit the investigation of the vocalised form **beze** to the modern non-fictional texts in the corpus, i.e. to press and EU legislation. All collocational types with the form **beze** exhibiting more than one occurrence are listed in the following table.

beze zbytku	259	100%	bez zbytku	1
beze změn	163	97%	bez změn	5
beze změny	206	91%	bez změny	21
beze slov	57		bez slov	0
beze slova	59		bez slova	2
beze stop	23		bez stop	0
beze stopy	92	84%	bez stopy	17
beze strachu	35		bez strachu	14
beze mě	18		bez mě	0
beze mne	22		bez mne	0
beze zbraně	18		bez zbraně	1
beze zbraní	15		bez zbraní	1
beze ztrát	13		bez ztrát	10
beze ztráty	18	9%	bez ztráty	193
beze studu	18		bez studu	7
beze smyslu	15		bez smyslu	7
beze spěchu	16		bez spěchu	1
beze svědků	16		bez svědků	2
beze jmen	5		bez jmen	1
beze jména	12		bez jména	10
beze všeho	11		bez všeho	0
beze všech	9		bez všech	10
beze vší	5		bez vší	0
beze stínu	11		bez stínu	9
beze sporu	9		bez sporu	1
beze spánku	8		bez spánku	7
beze sněhu	6		bez sněhu	35
beze světla	6		bez světla	7

beze švů	5	bez švů	1
beze smlouvy	4	bez smlouvy	65
beze srážek	3	bez srážek	3
beze stravy	3	bez stravy	1
beze škody	3	bez škody / Škody	4
beze cla	2	bez cla	34
beze cti	2	bez cti	0
beze strastí	2	bez strastí	0
beze snahy	2	bez snahy	10
beze svíčky	2	bez svíčky	0

Table 91. Vocalisation of the preposition **bez** in the contemporary part of the Czech corpus.

Besides the above collocations, 27 single occurrences were still encountered. Sibilant initial /CCCV/: *beze skvrny, beze stran, beze stromů, beze stropu, beze střech, beze svědomí, beze zpráv;* /CCV/: *beze shonu, beze schodků, beze schodů, beze slunce, beze slůvka, beze smíchu, beze smutku, beze spár, beze stínu, beze škod, beze zbytu, beze zdi, beze zloby, beze zmatků, beze zmínky, beze značky, beze známek, beze známky;* /CV/: *beze sedla*. Only *beze jejích* does not have an initial sibilant. Hence, in these texts there is evidently a certain freedom in terms of individual choice of form, but this freedom exists almost exclusively before sibilant initial consonant clusters.

The only collocational types with seemingly obligatory vocalisation are **beze mne**, **beze mě** *without me*, with the 1st person personal pronoun spelled in two possible ways, **beze slov** *without words*, **beze stop** *without trace*, **beze zbytku** *completely* (lit. *without remainder*).

Very high relative frequencies are also displayed by the collocations **beze slova** (59/2) *without a word*, **beze stopy** (92/17) *without a trace*, **beze změn** (163/5) *without changes*, **beze změní** (206/21) *without change*, **beze zbraně** (18/1) *without weapon*, **beze zbraní** (15/1) *without weapons*, **beze spěchu** (10/1) *without hurry*, **beze sporu** (9/1) *without dispute*, **beze svědků** (16/2) *without witnesses*.

A medium relative frequency is displayed by the collocations **beze jména** (12/10) *without name*, **beze studu** (18/7) *without shame*, **beze strachu** (25/14) *without fear*, **beze smyslu** (15/7) *without meaning*, **beze ztrát** (13/10) *without losses*, **beze spánku** (8/7) *without sleep*, **beze srážek** (3/3) *without rainfall*, **beze stínu** (11/9) *without shadow*, **beze světla** (6/7) *without light*, **beze škody** (3/4) *without harm*, whereas a low relative frequency is displayed by the collocations **beze stráty** (18/193) *without loss*, **beze škod** (1/3) *without damages*, **beze snahy** (2/10) *without effort*, **beze sněhu** (6/35) *without snow*, **beze smlouvy** (4/65) *without contract*, **beze cla** (2/34) *without customs duty*.

A high relative frequency is also displayed by a few quite infrequent collocational types, hence it is difficult to determine the status of **beze švů**

(5/1) *without seams*, **beze jmén** (5/1) *without names*, **beze stravy** (3/1) *without food*, **beze všeho** (11/0) *without all, without doubt*, **beze vši** (5/0) *without all*, **beze cti** (2/0) *without honour*, **beze strasti** (2/0) *without sorrows* and **beze svíčky** (2/0) *without the candle*.

It seems that apart from a handful of very frequent collocations, no apparent patterns can be traced in the above data. Rather, a certain degree of free variation seems to be present before words beginning with a sibilant-initial consonant cluster. This assumption is supported by the data in the following table, which displays the frequency of vocalisation before words beginning with a sibilant followed by a vowel or a consonant respectively.

beze SV	1	0%	bez SV	3425
beze SC	1096	24%	bez SC	3454

Table 92. Vocalisation of the preposition **bez** before an initial sibilant in the contemporary part of the Czech corpus.

8.1.1.2. nad(e)

For the preposition **nad(e)**, the range of vocalisations occurring in the corpus is far more restricted.

nade d*	31	2%	nad d*	1,504
nade h*	17	1%	nad h*	2,005
nade k*	3	0%	nad k*	1,790
nade l*	1	0%	nad l*	3,790
nade m*	204	9%	nad m*	2,009
nade n*	4	0%	nad n*	3,803
nade p*	6	0%	nad p*	3,089
nade r*	5	0%	nad r*	1,681
nade s*	9	0%	nad s*	4,201
nade t*	5	0%	nad t*	3,570
nade v*	340	12%	nad v*	2,443
nade z*	6	0%	nad z*	1,240

Table 93. Vocalisation of the preposition **nad** in the Czech corpus.

Furthermore, out of the 629 vocalisations, only 218 occur in the contemporary part of the corpus. The corresponding number of non-vocalised form is 42,136 in the whole corpus and 36,697 in the contemporary part, which means that the relative rate of vocalisation is 1.5% in the whole corpus, but only 0.6% in the contemporary part of the corpus.

The following collocational types are present in the contemporary part of the corpus.

nade dnem	1	nad dnem	3
nade dveře	1	nad dveře	1
nade dveřmi	6	nad dveřmi	4
nade mne	3	nad mne	0
nade mnou	46	nad mnou	0
nade Mží	1	nad Mží	0
nade vše	71	nad vše	1
nade všechn*	21	nad všechn*	23
nade všemi	12	nad všemi	46
nade vši	16	nad vši	0
nade vší	23	nad vší	3
nade vším	17	nad vším	30

Table 94. Vocalisation of the preposition **nad** in the contemporary part of the Czech corpus.

As could be expected, only the vocalised form of the preposition occurs with the 1st person singular person pronoun.

A second frequent context is the different inflectional forms of the pronoun **všechn** *all*. The collocational type **nade vše** is used as a phrase with the abstract meaning *above all*, and in this meaning vocalisation seems to be obligatory. The only counterexample occurs in the longer phrase **nad všechny a nad vše**, with the more literal meaning *over everybody and over everything*. The collocations **nade vši** and **nade vší** only occur with the following head word **pochybnost** meaning *beyond all doubt*, **vši** actually being a frequent spelling mistake. The vocalised form seems to be obligatory in this context too. The collocations **nad(e) všemi**, **nad(e) všechn*** *over all*, with the vocalised form of the preposition, are seldom followed by a nominal head word, whereas the collocations with the non-vocalised form are predominantly followed by a nominal head word. For **nad(e) vším** no evident pattern can be observed.

The status of **nade dnem** (1/3) *above the bottom*, **nade dveře** (1/1) *above the door* (direction), **nade dveřmi** (6/4) *above the door* (position) and **nad(e) Mží** (1/0) *above (the river) Mže* is difficult to evaluate because of the low number of instances in the corpus.

8.1.1.3. od(e)

Again, for the preposition **od(e)** *from*, a whole range of different degrees of vocalisation is displayed, depending on the following context.

ode b*	2	0%	od b*	3,131
ode č	1	0%	od č*	3,791
ode d*	1,364	17%	od d*	6,548

ode ch*	1	0%	od ch*	673
ode k*	1	0%	od k*	5,656
ode m*	747	12%	od m*	5,627
ode r*	9	0%	od r*	10,500
ode s*	4	0%	od s*	14,014
ode t*	1	0%	od t*	8,882
ode v*	144	2%	od v*	6,580
ode z*	38	0%	od z*	8,028
ode 2	2		od 2	52

Table 95. Vocalisation of the preposition **od** in the Czech corpus.

Out of a total of 2,314 vocalisations, only 1,567 occur in the contemporary part of the corpus. The total number of non-vocalised occurrences are 145,907, of which 132,677 occur in the contemporary texts, reducing the relative rate of vocalisation from 1.6% in the whole corpus to 0.6% in the contemporary part of the corpus.

The following collocational types are found in the contemporary part of the corpus.

ode dna	68		od dna	8
ode dne	245	99%	od dne	2
ode dneška	851	97%	od dneška	24
ode dnešního	3	1%	od dnešního	282
ode dnů	1		od dnů	0
ode dveří	20		od dveří	27
ode mě	106	99%	od mě	1
ode mne	218	99%	od mne	2
ode mně	4		od mně	1
ode včerejšího	1	0%	od včerejšího	205
ode všeho	11		od všeho	54
ode všech	7	3%	od všech	197
ode zdi	32		od zdi	23

Table 96. Vocalisation of the preposition **od** in the contemporary part of the Czech corpus.

Quite unexpectedly, four occurrences were encountered with the non-vocalised form of the preposition before the 1st person personal pronoun. However, on closer inspection, only two instances are unquestionable. One instance is found in an ungrammatical clause, and another instance contains **od mně blízke osoby** *from a person close to me*, where the preposition does not govern the pronoun, but rather the nominal phrase **blízke osoby**; the pronoun is a further determiner of the adjective **blízke**. However, we still have one

instance each of **od mě** and **od mne** which cannot be explained otherwise than as writing mistakes or effects of the authors' individual choices.

Further, it can be observed that vocalisation is close to obligatory in the collocational type **ode dne** (245/2) *from the day*. Highly frequent are also **ode dneška** (851/24) *from today* and **ode dna** (68/8) *from the bottom*. Rather infrequent are the two collocational types **ode všeho** (11/54) and **ode všech** (7/197). No distributional patterns can be observed for any of these five collocational types, and a certain freedom of choice seems to exist.

The collocational type **ode zdi** (32/23) *from the wall* occurs 29 times in the longer phrase **ode zdi ke zdi** *from wall to wall*. The same phrase does not occur with the non-vocalised form of the preposition, whereas the remaining three occurrences with the vocalised form are of the same general semantic character as the 23 occurrences with the non-vocalised form. A similar situation is found with regard to the collocational type **ode dveří** (20/27) *from the door*. It occurs 5 times in the longer phrase **ode dveří ke dveřím** *from door to door*, but the non-vocalised form **od** also occurs once in the very same phrase. For the remaining occurrences, no distributional patterns could be observed.

Finally, quite marginal occurrences were found with two collocational types containing derived adjectives **ode dnešního** (3/282) *from today's* and **ode včerejšího** (1/205) *from yesterday's*, as well as the collocation **ode dnů** (1/0) *from the days*, which is highly interesting because of its relationship to the highly frequent collocational type **ode dne** *from the day*.

8.1.1.4. pod(e)

For the preposition **pod(e)** *under*, there is a wide range of contexts exhibiting vocalisation in the corpus.

pod e c*	1	0%	pod c*	296
pod e d*	7	0%	pod d*	1,669
pod e h*	2	0%	pod h*	1,791
pod e j*	19	1%	pod j*	1,862
pod e k*	1	0%	pod k*	2,728
pod e m*	58	6%	pod m*	964
pod e n*	1	0%	pod n*	4,617
pod e p*	9	0%	pod p*	4,761
pod e ř*	2		pod ř*	59
pod e s*	8	0%	pod s*	3,431
pod e š*	1	0%	pod š*	510
pod e v*	5	0%	pod v*	3,850
pod e z*	10	1%	pod z*	1,461

Table 97. Vocalisation of the preposition **pod** in the Czech corpus.

However, out of the 124 instances of **pode**, only 40 occur in the contemporary part of the corpus, and out of these, as many as 23 are due to spelling mistakes. It occurs six times instead of an intended **podle**, e.g. **pode jeho slov** *according to his words*, and 17 times it represents the homonymous prefix which has been written separated from the following root, e.g. **pode psal** *signed*. This leaves only 17 instances of the vocalised form of the preposition **pod(e)** in the contemporary part of the corpus.

The corresponding number of occurrences of **pod** is reduced from 37,168 in the total corpus to 32,022 in the contemporary parts, which leads to a reduction of the relative rate of vocalisation from 0.3% to 0.05%.

Only two collocational types with **pode** occur in the contemporary part of the corpus.

pode dnem	2	7
pode mnou	15	0

Table 98. Vocalisation of the preposition **pod** in the contemporary part of the Czech corpus.

As expected, only the vocalised form of the preposition occurs in collocations with the 1st person singular personal pronoun. Further, both collocations with the word **dnem** are instances of the lexeme **dno** *bottom* occurring after expressions of distance consisting of numerals followed by a measuring unit, e.g. **4500 metrů pode dnem** *4500 metres below the bottom*. Out of the seven occurrences with the non-vocalised form, only one occurrence is of this type.

8.1.1.5. před(e)

Once again, there is a whole range of vocalisations present in the older parts of the corpus. 28 instances of the homonymous verb **přede** *spins; purrs* are also present in the corpus, as well as six instances of the prefix **přede** (three written with a hyphen and three separately). These 34 occurrences have been disregarded in the following table.

přede b*	1	0%	před b*	2,456
přede d*	278	3%	před d*	8,700
přede j*	1	0%	před j*	2,106
přede k*	1	0%	před k*	6,683
přede m*	503	15%	před m*	2,798
přede s*	4	0%	před s*	10,889
přede v*	127	2%	před v*	8,196
přede ž*	3	2%	před ž*	155

Table 99. Vocalisation of the preposition **před** in the Czech corpus.

Out of the 918 vocalisations, 491 occur in the contemporary part of the corpus. The corresponding numbers of non-vocalised occurrences are 99,877 and 92,278 respectively, reducing the relative rate of vocalisation from 0.9% overall, to 0.5% in the modern part of the corpus.

All collocations with the vocalised form **přede** occurring in the modern part of the corpus are presented in the following table.

přede dnem	20		před dnem	28
přede dvěma	123	3%	před dvěma	3,559
přede dveře	3		před dveře	15
přede dveřmi	44		před dveřmi	39
přede mě	6		před mě	0
přede mne	12		před mne	0
přede mnou	217	100%	před mnou	0
přede všemi	25		před všemi	43
přede vším	41		před vším	7

Table 100. Vocalisation of the preposition **před** in the contemporary part of the Czech corpus.

Once again, the use of the vocalised form of the preposition proves to be obligatory before the 1st person personal pronoun. For the collocational types involving **dnem** *day*, **dvěma** *two*, **dveře** *door*, no patterns can be traced, hence the variation seems to be free, with almost equal frequencies of both forms of the preposition before **dnem** *day*, although there is a clear domination of **před** before **dvěma** and **dveře** (denoting direction). Before the form **dveřmi** (denoting position), both forms of the preposition are equally common and may be used in the general spatial sense of the word. However, in a figurative temporal meaning *just around the corner*, only the vocalised form of the preposition occurs in the corpus data, e.g. **Členství v NATO je přede dveřmi**. *Membership of NATO is just around the corner*.

Finally, when it comes to the instances of this preposition followed by the pronominal forms **vším** *all* (sg.) and **všemi** *all* (pl.), quite clear patterns can be observed. Before **vším**, the vocalised form is only used when the collocation has the somewhat figurative meaning *above all*, i.e. *more than anything else*, whereas in the literal spatial meaning, only the non-vocalised form occurs. The collocations with the vocalised form do not actually comply with modern Czech orthography, as in this meaning it should be written as one word, i.e. **přede vším**, which occurs 29,461 times in the modern part of the corpus. The situation is quite different before **všemi**, where the collocational type **přede všemi** *in front of all, before everybody* with vocalised form of the preposition is used independently, without a following noun phrase 21 times, and with a following noun phrase only 4 times, whereas the collocational type with the

non-vocalised form of the preposition is used only 8 times independently, but as many as 35 times followed by a noun phrase. This means that the vocalised form is used in 72% of the instances when followed by **všemi** used independently, but only in 10% of the instances where the following **všemi** is a determiner of a following head noun.

8.1.1.6. př(es)e

The spelling of this preposition deviates from the otherwise phonological principles of Czech orthography. A more adequate spelling would have been **přez(e)**, and the vocalised form is always pronounced [přɛzɛ].

With this preposition, there are fewer different collocations present in the older parts of the corpus with the vocalised form of the preposition.

přese m*	9	0%	přes m*	1,358
přese r*	1	0%	přes r*	1,074
přese s*	3	0%	přes s*	2,999
přese v*	106	4%	přes v*	2,366

Table 101. Vocalisation of the preposition **přes** in the Czech corpus.

Out of the 119 vocalisations, however, only 45 occur in the contemporary part of the corpus. The corresponding numbers of non-vocalised occurrences are 33,335 and 30,233 respectively, reducing the relative rate of vocalisation from 0.4% in the whole corpus to 0.1% in the modern part of the corpus.

All collocations with the vocalised form of the preposition **př(es)e** occurring in the contemporary part of the corpus are presented in the following table.

přese mě	4		přes mě	4
přese mne	1		přes mne	0
přese vše	1		přes vše	2
přese všechna	2		přes všechna	77
přese všechno	32		přes všechno	46
přese všechnu	2		přes všechnu	69
přese všechny	3	1%	přes všechny	392

Table 102. Vocalisation of the preposition **přes** in the contemporary part of the Czech corpus.

Most unexpectedly, both forms of the preposition are equally frequent before the 1st person singular personal pronoun **mě**, and no systematic difference in usage can be noticed. The number of instances is too small, however, to draw any significant conclusions.

Before the different forms of the pronoun **všechen** *all*, two strong

tendencies can be observed. First, the vocalised form only occurs in the circumstantial meaning *despite*, e.g. **přese všechny výhrady** *despite all reservations*, whereas the non-vocalised form occurs both in this circumstantial meaning and in the spatial meaning *through, across, over*, e.g. **přes všechna patra luxusního domu** *throughout every floor of the luxurious building*. This might however be due to the fact that the circumstantial meaning is by far the most frequent in collocations with the pronoun **všechen**.

Furthermore, the vocalised form occurs before **všechno** as an independent head 31 times and twice before **vše** when this pronoun is not followed by a noun phrase, whereas it only occurs once before **všechno** as the determiner of a following noun phrase. For obvious reasons almost all other forms of this pronoun are followed by a head noun, and not surprisingly the number of instances with the vocalised form of the preposition is very low.

8.1.1.7. **zpod(e)**

The vocalised form of the preposition **zpod(e)** *from beneath* was not encountered in the corpus data.

8.1.1.8. **zpřed(e)**

The vocalised form of the preposition **zpřed(e)** *from in front of* was not encountered in the corpus data.

8.1.1.9. **poblíž(e)**

The vocalised form **poblíže** *close (by), not far (from)* occurs 38 times in the corpus, whereas the non-vocalised form **poblíž** occurs 2,013 times. This gives a frequency of the vocalised form of 2%. However, in the contemporary part of the corpus the numbers are 1/1949, which suggests that the vocalised form may well be rather archaic. The instances of the vocalised form of the preposition do not exhibit any pattern with regard to the initial phoneme of the following word. Quite unexpectedly, it also occurs a handful of times before words beginning with a vowel.

Furthermore, apart from being used as a preposition, e.g. **poblíže italské Spezie** *not far from Italian Spezia*, this lexeme may also be used independently as a spatial adverb, e.g. **poblíže ležely dvě nakupeniny balvanů** *close by were two piles of large stones*. The vocalised form is encountered in the corpus 36 times as a preposition and twice as an adverb. Without any more detailed investigation, it is apparent that the distribution of the non-vocalised form between these two parts of speech is quite similar.

The same alternation as in **poblíže** is also found in a number of comparative adverbs, such as **výš(e)** *higher*, **níž(e)** *lower*, **blíž(e)** *closer*, **lépe/líp** *better*, **víc(e)** *more*, **méně/míň** *less* etc., of which one or the other form is preferred depending on stylistic register (Sgall & Hronek 1992: 45).

8.1.1.10. skrz(e)

For this preposition, the vocalised form, with 926 instances, is far more frequent than the non-vocalised form with 323 instances. Once again, no patterns in the distribution of the two forms can be identified based on the initial phoneme of the word following the preposition. There are again many instances with a following word beginning with a vowel, e.g. **skrze alianci** *through the alliance*.

Instead, another pattern immediately becomes quite clear, namely that the vocalised form is highly preferred in a figurative instrumental meaning, as in the example above or in **se na vše dívají skrze brýle peněz a obchodu** *they look at everything through the lenses of money and business*, whereas the non-vocalised form is preferred in the concrete spatial meaning *through*, e.g. **skrz zdi** *through walls*. Only a handful of instances with the vocalised form in a very concrete spatial meaning are encountered, mainly in the older parts of the corpus, e.g. **skrze dvěře** *through the door*, used twice by Božena Němcová.

8.1.1.11. dík(y)

When it comes to the preposition **dík(y)** *thanks to*, there are 11,860 instances of the form **díky** and 737 instances of **dík** in the corpus, which would give a relative frequency of 94% for the vocalised form **díky**. Hence, the vocalised form is evidently unmarked, whereas the non-vocalised form is marked.

However, a substantial complication is that **dík** is frequently used as an abbreviation for a certain source in the newspaper *Mladá fronta*. Furthermore, the homonymous noun **dík** *thank*, e.g. **za to velký dík** *a huge thank for that*, occurs quite frequently. Excluding all abbreviations and all instances of the noun, as well as all instances from the older part of the corpus, we only retain 145 instances representing the non-vocalised form of the preposition in the contemporary part of the corpus. Among the 11,659 occurrences of the form **díky** in the contemporary part of the corpus, only about 3% do not represent the preposition, but rather the plural of the noun.⁹³ The relative frequency of the vocalised form of the preposition thereby rises to 98.7%.

No pattern could be observed in the distribution of the marked non-vocalised form of this preposition. It can however be observed that in the older part of the corpus, the relative frequency of the prepositional usage is much lower. In 19th century texts, the noun occurs far more frequently than the preposition in both its forms.

8.1.1.12. krom(ě)

For the preposition **krom(ě)** *besides, apart from*, there are 17,652 instances of the

⁹³ Based on the analysis of 100 random instances.

vocalised form and 776 instances of the non-vocalised form in the corpus, which gives a relative frequency of 96% for the form **kromě**. Hence, once again, the vocalised form is evidently unmarked, whereas the non-vocalised form is marked.

However, quite interestingly, there has occurred a clearly noticeable change in usage over time. In the older part of the corpus, this preposition is considerably less frequent. The non-vocalised form occurs 428 times, whereas the vocalised form occurs only 378 times, which gives a relative frequency of only 47% for the vocalised form **kromě**. In the contemporary part of the corpus the corresponding figures are 348 instances of the non-vocalised form and 17,274 instances of the vocalised form, giving a relative frequency of 98.0% for the vocalised form.

No pattern could be observed in the distribution of the marked non-vocalised form of this preposition. It seems plausible, however, with regard to the data from the older part of the corpus, that the non-vocalised form could represent a somewhat archaic, bookish or stylistically refined usage.

8.1.1.13. **navzdor(y)**

Once again, the same pattern is repeated for the preposition **navzdor(y)** *despite*. There are 2,727 instances of the vocalised form and 83 instances of the non-vocalised form in the corpus, which gives a relative frequency of 97.0% for the form **navzdory**, hence the vocalised form is evidently unmarked.

For this preposition again, there has occurred a clearly noticeable change in usage over time. In the older part of the corpus, this preposition is considerably less frequent. The non-vocalised form occurs 78 times, whereas the vocalised form occurs only 33 times, which gives a relative frequency of only 30% for the vocalised form **navzdory**. In the contemporary part of the corpus the corresponding figures are 5 instances of the non-vocalised form and 2,694 instances of the vocalised form, giving a relative frequency of 99.8% for the vocalised form.

No formal pattern can be observed in the distribution of the marked non-vocalised form of this preposition. It seems quite plausible, however, with regard to the data from the older part of the corpus, that the non-vocalised form could represent a somewhat archaic, bookish or stylistically refined usage.

8.1.2. **Asyllabic Prepositions**

Just as in the case of Polish and Slovak, the working hypothesis is that the Czech asyllabic prepositions are likely to occur in their vocalised form before phonological strings together with which the purely consonantal prepositions would otherwise give rise to phonological structures that are disallowed by the phonological principles of Czech.

An additional complication with regard to the Czech data is the homonymy of the vocalised preposition **se** *with* and the reflexive pronoun **se** *oneself*. Because of this homonymy, a complete investigation of the preposition **s(e)** is beyond the scope of this work. Instead, the findings regarding the preposition **z(e)**, which on phonological grounds might be expected behave quite similarly, is taken as a point of departure, and the behaviour of these two prepositions is then compared, as they would be expected to behave much in the same way.

During the investigation of the remaining three asyllabic prepositions in Czech, it turns out that the Czech corpus, like the Slovak corpus, contains quite a lot of problematic instances of **k**, **v**, **z** which are not occurrences of these prepositions, but constitute initials, parts of abbreviations, or parts of headlines printed with spaces between each letter. In order to exclude this type of disruptive data from the concordance listing, the search strings have been narrowed so that the programme retrieves only instances of the non-vocalised prepositions **k**, **v**, **z** followed either by a space and at least two letters of the alphabet or by a space and a digit. Furthermore, all instances of **k**, **v**, **z** directly followed by a full stop are left out of the search results.

8.1.2.1. **k(e)/ku**

All instances of the vocalised form of the preposition **k(e)/ku** were extracted and sorted with respect to the first grapheme of the following word. Then all occurrences of the non-vocalised form **k** occurring in the same contexts were counted for comparison.

During this process, some instances were encountered that did not represent instances of the preposition. **Ke** and **ku** occurred 20 times in names, mainly of East-Asian origin. Wrongly inserted spaces or hyphens occurred 521 times, e.g. **Mi ke** *Mike*, **ro-ku** *year*. The two forms also occurred three times as abbreviations with unclear meaning. Further, the forms **ke/ku** occurred 104 times before an initial **c** or **z** not representing their standard Czech sound values [ts] and [z] respectively, e.g. **ku Clintonovi** *to Clinton*, as well as twice before Roman numerals beginning with an **I**. **Ku** also occurred once as an interjection and four times in sentences which were in Slovak. Finally there were three uninterpretable occurrences. All these occurrences of the vocalised form of the preposition have been omitted from the data in the following table, in order not to distort the analysis.

For the non-vocalised form of the preposition, there were too many abbreviations to be able to give a survey of them. They were all excluded from the data. Once the letter **k** itself is mentioned, and once **k** is used as a label of a coefficient. There were also a few writing mistakes like **v k kladném** instead of **v kladném** *in the positive*. Furthermore, there were 15 instances where the preposition **k** occurred before words of foreign origin beginning with a (g) which is normally not pronounced as [g], e.g. **k gentlemanům** *to(wards) the*

gentlemen or **k know-how** to(wards) know-how. All of these instances have been disregarded in the following table.

ke b*	75	2%	ku b*	37	1%	k b*	4,421
ke c*	208	7%	ku c*	12	0%	k c*	2,920
ke č*	462	11%	ku č*	27	1%	k č*	3,701
ke d*	1,404	7%	ku d*	61	0%	k d*	17,896
ke e*	0	0%	ku e*	1	0%	k e*	1,838
ke f*	3	0%	ku f*	0	0%	k f*	2,086
ke g*	469	100%	ku g*	0	0%	k g*	1
ke h*	332	7%	ku h*	26	1%	k h*	4,227
ke j*	71	1%	ku j*	96	1%	k j*	8,949
ke k*	10,407	97%	ku k*	291	3%	k k*	23
ke l*	35	1%	ku l*	15	0%	k l*	4,442
ke m*	1,118	12%	ku m*	25	0%	k m*	8,485
ke n*	2	0%	ku n*	7	0%	k n*	30,994
ke o*	0	0%	ku o*	16	0%	k o*	12,808
ke p*	32	0%	ku p*	660	2%	k p*	28,833
ke q*	3		ku q*	0		k q*	0
ke r*	41	0%	ku r*	8	0%	k r*	8,208
ke ř*	2	0%	ku ř*	1	0%	k ř*	1,093
ke s*	13,630	54%	ku s*	102	0%	k s*	11,461
ke š*	752	44%	ku š*	23	1%	k š*	945
ke t*	397	1%	ku t*	31	0%	k t*	31,259
ke v*	2,304	10%	ku v*	40	0%	k v*	20,269
ke x*	9		ku x*	0		k x*	0
ke z*	5,836	36%	ku z*	30	0%	k z*	10,188
ke ž*	19	1%	ku ž*	4	0%	k ž*	2,012
ke 0*	0		ku 0*	2		k 0*	0
ke 1*	130	78%	ku 1*	37	22%	k 1*	0
ke 2*	39		ku 2*	23		k 2*	0
ke 3*	59		ku 3*	15		k 3*	0
ke 4*	30		ku 4*	18		k 4*	0
ke 5*	0		ku 5*	5		k 5*	0
ke 6*	0		ku 6*	2		k 6*	0
ke 7*	1		ku 7*	6		k 7*	0
ke 8*	0		ku 8*	2		k 8*	0
ke 9*	0		ku 9*	6		k 9*	0

Table 103. Vocalisation of the preposition **k** in the Czech corpus.

Already from the above data it becomes quite clear that there are certain

contexts which favour the usage of the vocalised forms of the preposition. To a high degree, these contexts also resemble the contexts encountered in Slovak.

The highest frequencies are found before words beginning with ⟨g⟩ and ⟨k⟩, as well as with ⟨x⟩, ⟨q⟩ and all the digits. There are also quite high frequencies before ⟨s⟩, ⟨š⟩, ⟨z⟩, as well as significant frequencies before ⟨m⟩, ⟨c⟩, ⟨č⟩, ⟨h⟩.

The distribution between the two different vocalised forms is also quite intriguing, and this preposition represents a totally unique case by exhibiting three different shapes.

However, as has already been noticed on a number of occasions, a closer inspection of the data immediately shows that there are considerable differences in the usage of the different forms of prepositions between the older parts of the corpora and the contemporary parts. Therefore, in order to make the data as clear as possible from the point of view of present day usage, a comparison is first made between the older and the contemporary parts of the corpus. The findings are presented in the following table.

total	ke	37,907	14.1%	ku	1,631	0.6%	k	229,740
old	ke	3,960	10.6%	ku	974	2.6%	k	32,266
contemporary	ke	33,947	14.6%	ku	657	0.3%	k	197,474

Table 104. Vocalisation of the preposition **k** in the old and the contemporary part of the Czech corpus.

As can immediately be seen, the total rate of vocalisation has risen somewhat over time, from 13.2% in the older parts of the corpus to 14.9% in the contemporary parts of the corpus. Furthermore, the frequency of the form **ke** has risen even more, whereas the frequency of the form **ku** has fallen dramatically from 2.6% to merely 0.3%. Hence, the general use of the vocalised forms seems to have been somewhat more restricted a hundred years ago, whereas the alternative form **ku** was used much more frequently than today. Therefore, it can be suspected that the usage of **ku**, as well as the usage of **k**, where **ke** would be systemically expected in present day usage, may have a function as stylistic means with a bookish or archaic flavour.

Against this background we will examine only the data from the contemporary part of the corpus in greater detail, beginning with the seemingly obligatory usage of **ke** before words beginning with /k/ and /g/, then moving on to the contexts beginning with one of the sibilants /s/, /z/, /š/, /ž/, followed by a closer investigation of the seemingly obligatory usage of one of the vocalised forms before digits, and finishing with the remaining contexts, which trigger vocalisation less frequently.

When it comes to collocations with the preposition **k(e)/ku** followed by a word beginning with /k/ and /g/, i.e. also including the instances before ⟨x⟩ and ⟨q⟩, the vocalised form **ke** clearly seems to be obligatory. Only three

instances of **k k*** occur in the contemporary corpus, e.g. **postoj k komisi** *the attitude towards the commission*. One single instance of **k g*** occurs: **k genocidě**.

Collocations with **ku k*** occur seven times: **ku kamennému mostu** *towards the stone bridge*, **ku konci** (4x) *towards the end*, **ku kterémuž** *towards which*. The first and the last are found in citations from an older text, whereas the four instances of **ku konci** occur in normal contemporary texts. The corresponding **ke konci** however occurs 1,316 times.

The next interesting context with regard to vocalisation of this preposition is before words beginning with a sibilant. It can easily be observed that the vocalisation depends on the second phoneme of the following word. The rate of vocalisation is very high before words beginning with a sibilant initial consonant cluster and almost non-existent before words beginning with a sibilant followed by a vowel. The relevant data are presented in the following table.

ke sC*	12,902	88%	ku sC*	21	0%	k sC*	1,795
ke šC*	702	93%	ku šC*	7	1%	k šC*	47
ke zC*	7,989	95%	ku zC*	4	0%	k zC*	454
ke žC*	14	79%	ku žC*	0	0%	k žC*	4
ke sV*	2	0%	ku sV*	4	0%	k sV*	6,013
ke šV*	1	0%	ku šV*	12	2%	k šV*	676
ke zV*	19	0%	ku zV*	0	0%	k zV*	8,522
ke žV*	0	0%	ku žV*	0	0%	k žV*	1,632

Table 105. Vocalisation of the preposition **k** before sibilants in the contemporary part of the Czech corpus.

From the above data, it becomes clear that the vocalised form of this preposition **ke** is practically not used before a following word beginning with a sibilant followed by a vowel.

All of the 16 instances of the form **ku** in the context of a following word beginning with sV*/šV* occur with a following number, e.g. **ku sedmi** *to seven*, **ku šestnácti** *to sixteen*. In the remaining contexts as well, **ku** is quite often followed by a number, e.g. **ku sto jedné** *to one hundred and one*. This fact is well in line with the data in Table 103, where digits are never preceded by the non-vocalised form of this preposition, and most digits are more often preceded by the form **ku** than the form **ke**. We will return to the issue of the distribution of the two vocalised forms before numerals shortly, but first we will look a little more deeply into the distribution of the preposition before words beginning with a sibilant followed by a consonant, in order to investigate whether the type of consonant following the sibilant might be the deciding factor. The relevant data are presented below.

ke Sč*	3	100%	ku Sč*	0		k Sč*	0
ke Sk*	1253	99%	ku Sk*	6	0.5%	k Sk*	8
ke Sc*	479	98%	ku Sc*	0		k Sc*	9
ke Sp*	2699	96%	ku Sp*	3	0%	k Sp*	107
ke Sd*	707	96%	ku Sd*	2	0%	k Sd*	30
ke Sv*	4541	93%	ku Sv*	8	0%	k Sv*	343
ke St*	4069	91%	ku St*	6	0%	k St*	387
ke Sh*	290	91%	ku Sh*	0		k Sh*	29
ke Sř*	74	88%	ku Sř*	0		k Sř*	10
ke Sb*	248	87%	ku Sb*	1	0%	k Sb*	35
ke Sj*	227	85%	ku Sj*	0		k Sj*	39
ke Sl*	1605	85%	ku Sl*	4	0%	k Sl*	278
ke Sn*	1081	84%	ku Sn*	2	0%	k Sn*	208
ke Sf*	19	79%	ku Sf*	0		k Sf*	4
ke Sm*	1518	72.5%	ku Sm*	0		k Sm*	575
ke Sr*	442	65%	ku Sr*	0		k Sr*	235

Table 106. Vocalisation of the preposition **k** before sibilant-initial consonant groups in the contemporary part of the Czech corpus.

There are no dramatic differences between different types of consonants in the above table, but slight tendencies can be observed with regard to the type of consonant following the sibilant. In the table below, the contexts have been grouped into sibilant + sonorant (SSo), sibilant + fricative obstruent (SFr), and sibilant + plosive/affricative obstruent (SPI) respectively.

ke SPI*	9458	94%	ku SPI*	18	0%	k SPI*	576
ke SFr*	4924	93%	ku SFr*	8	0%	k SFr*	386
ke SSo*	4873	78%	ku SSo*	6	0%	k SSo*	1335

Table 107. Vocalisation of the preposition **k** before sibilant-initial consonant groups in the contemporary part of the Czech corpus.

The difference is only significant between consonant clusters containing a sonorant consonant, as opposed to a non-sonorant consonant as the second element. However, another factor is probably important here, namely the possibility of the sonorants /l/ and /r/ occurring in the syllabic position, normally occupied by a vowel, i.e. filling the function of a syllable nucleus. An investigation of the different types of /r/ in the context investigated suggests that the rate of vocalisation before a sibilant followed by a “consonantal” /r/ constituting part of the onset (Sr_c) is just as high as before any other combination of sibilant + consonant, whereas vocalisation before a sibilant followed by a “vocalic” /r/, constituting a syllable nucleus (Sr_v), is practically non-existent.

ke Sr _c * 441 94%	ku Sr _c * 0	k Sr _c * 29
ke Sr _v * 1 0.5%	ku Sr _v * 0	k Sr _v * 206

Table 108. Vocalisation of the preposition **k** before a sibilant followed by different types of /r/ in the contemporary part of the Czech corpus.

The next prominent context triggering vocalisation of the preposition **k** is that of numerals. First of all, it is striking that not one single instance of the non-vocalised form is found before a digit. Secondly, it is also striking that the frequency of the vocalised form **ku** is significantly higher before digits than elsewhere. The explanation, however, is quite easy to find, on closer inspection of the relevant data. It turns out that all occurrences of the vocalised form **ku** before digits are found in a context where the preposition is used to express a relationship between two numbers, e.g. **jasnou většinou 81 ku 16 hlasům** *with the clear majority of 81 to 16 votes*. On the other hand, practically all occurrences of the vocalised form **ke** are found before digits which, when they are pronounced, begin with 2, 3, 4 or 100. Only one single instance of a relationship was encountered, i.e. **Zeman prohrál rozdílem 8 ke 12 hlasům** *Zeman lost with a difference of 8 to 12 votes*. All the rest occur in other types of constructions, e.g. **ke 100. výročí** *on the occasion of the 100th anniversary*, **odsouzena ke 279 dnům vězení** *sentenced to 279 days in prison*, **nepatří ke 4500 členům** *doesn't belong to the 4500 members*. As can be seen, the digits involved are both cardinal and ordinal, and they occur in many different combinations, including 12, 13, 14, 20, 30, 40, 100, 200, 300, 400 etc.

As no single instance of the non-vocalised form of the preposition is found before digits, the corresponding contexts spelled out with letters were investigated. The data are presented below.

ke jedn* 0 0%	ku jedn* 84 5%	k jedn* 1559
ke prvn* 0 0%	ku prvn* 0 0%	k prvn* 672
ke druh* 179 40%	ku druh* 0 0%	k druh* 274
ke dvě* 291 88%	ku dvě* 16 5%	k dvě* 22
ke dvo* 38 28%	ku dvo* 0 0%	k dvo* 97
ke tři* 255 78%	ku tři* 20 9%	k tři* 51
ke tři* 66 61%	ku tři* 0 0%	k tři* 42
ke čtrn* 23	ku čtrn* 2	k čtrn* 0
ke čtvrt* 1	ku čtvrt* 0	k čtvrt* 0
ke čtyř* 179 85%	ku čtyř* 17 8%	k čtyř* 15
ke pět* 0	ku pět* 1	k pět* 65
ke pát* 0	ku pát* 0	k pát* 91
ke pad* 0	ku pad* 8 4%	k pad* 208
ke šede* 0	ku šede* 1	k šede* 33

ke šest*	0	ku šest*	11	6%	k šest*	177
ke sedm*	0	ku sedm*	4	3%	k sedm*	133
ke osm*	0	ku osm*	8	6%	k osm*	134
ke dev*t*	0	ku dev*t*	5		k dev*t*	54
ke des*t*	0	ku des*t*	12	7%	k des*t*	161

Table 109. Vocalisation of the preposition **k** before numerals not written as digits in the contemporary part of the Czech corpus.

First of all, it should be pointed out that in the above data, once again, every single instance of the form **ku** occurs in the aforementioned context of a comparative relationship between two figures, e.g. **devadesát procent Albánců ku deseti procentům Srbů** *ninety percent of Albanians versus ten percent of Serbs*, hence this is evidently an important principle, and the data clearly show that in this function, there is a close to exclusive preference for the form **ku**, which would suggest that there might be a good reason for treating **ku** in this usage as a separate lexeme, having only the form **ku**. This would then be similar to the separate lexeme **zo** *about, approximately* in Slovak. The special status of this form is further emphasised by the fact that it is also applied before words beginning with a vowel, such as **tři ku osmi** *three versus eight*.

Further, it becomes clear that the total lack of non-vocalised forms before digits, which was noted earlier, is somewhat accidental. When written with letters, practically all numerals occur with the non-vocalised form **k** to some extent. It can be observed that the vocalised form **ke** is preferred before figures beginning with 2, 3 and 4, whereas the non-vocalised form is the only form occurring before the rest of the basic numerals. On the basis of the data in Table 106, regarding vocalisation before a consonant cluster beginning with a sibilant, it has already been shown that vocalisation is heavily preferred before the type of initial consonant cluster found in the numeral **sto** *hundred*, which in turn explains the quite large amount of vocalisation before the digit **1** in Table 103.

As it has by now become quite evident that before digits, the form **ku** is semantically different from the forms **k/ke**, this seems like a good occasion to investigate the instances of the form **ku** found in contexts other than before basic numerals.

ke bilionu	0	ku bilionu	2	k bilionu	0		
ke konci	1,316	100%	ku konci	4	0%	k konci	0
ke lásce	0	ku lásce	4	k lásce	19		
ke milionu	0	ku milionu	3	k milionu	7		
ke podivu	0	ku podivu	4	k podivu	2		
ke podpoře	0	0%	ku podpoře	2	1%	k podpoře	265

ke pomoci	0		ku pomoci	13		k pomoci	39
ke Praze	0		ku Praze	26		k Praze	53
ke prospěchu	4	1%	ku prospěchu	239	89%	k prospěchu	25
ke příkladu	0		ku příkladu	6		k příkladu	4
ke příležitosti	0		ku příležitosti	4		k příležitosti	17
ke slávě	50		ku slávě	3		k slávě	2
ke snížení	427	97%	ku snížení	2	0%	k snížení	13
ke	143	97%	ku	2	1%	k	2
spokojenosti			spokojenosti			spokojenost	
ke svému	1,358	94%	ku svému	3	0%	k svému	80
ke škodě	138	97%	ku škodě	4	3%	k škodě	0
ke tisíci	0		ku tisíci	2		k tisíci	2
ke vlastnímu	0	0%	ku vlastnímu	2	1%	k vlastnímu	179
ke zdraví	20		ku zdraví	2		k zdraví	0

Table 110. Vocalisation of the preposition **k** in collocations with **ku** not followed by a numeral in the contemporary part of the Czech corpus.

Excluded from the above table are the following single instances: *ku bohaté, ku brání, ku chorovodu, ku chvále, ku cti, ku čtení, ku Essenu, ku hladině, ku hlavnímu, ku hráčům, ku hřichu, ku kamennému, ku kterémuž, ku lži, ku marce, ku nule, ku ploše, ku počtu, ku podivení, k podzemní, ku poradě, ku potřebám, ku poučení, ku prohrám, ku půl, ku radosti, ku sblížení, ku slovanským, ku své, ku škodolibému, ku štěstí, ku Švejkovi, ku tanci, ku vdávání, ku všeobecnému, ku vývozu, ku zvelebení, ku zvyšování.*

A brief check of these collocations suggests that the use of the form **ku** seems in the majority of instances to have a rather ornamental function, and many of the contexts also bear other signs of a somewhat archaic or otherwise individual usage. This also follows as a quite natural effect from the aforementioned finding that the frequency of the form **ku** has fallen dramatically over the last hundred years. That fact alone implies that it could readily be associated with a somewhat archaic flavour.

On the other hand, there are also a handful of collocations, namely **ku marce, ku nule, ku milionu, ku bilionu, ku půl, ku tisíci**, which again are found in expressions of relative comparison, such as e.g. **hodnota je už stanovena v poměru 1:1 ku marce německé** *the value is already established in the ratio of 1:1 against the German mark.* In this type of context the form **ku** has already proven to be the standard choice in contemporary usage. There are, however, a small number of exceptional cases, e.g. **jedna k tisíci** *one in a thousand.* In order to try to find out how common the use of the form **k** is in this context, the contemporary part of the corpus was searched for instances of the numerals 1–5 followed by the preposition **k(u)**, which is in turn followed

by a numeral.

jedna ku NUM	77	jedna k NUM	4
dvě ku NUM	13	dvě k NUM	0
tři ku NUM	13	tři k NUM	1
čtyři ku NUM	6	čtyři k NUM	1
pět ku NUM	9	pět k NUM	0
Total	118	95%	6

Table 111. Distribution of the forms **ku** and **k** between numerals expressing a numeric relation in the contemporary part of the Czech corpus

The findings in the above table make it somewhat more complicated to establish a separate lexeme **ku** *versus*, compared to, as it will then also be necessary to postulate an alternative optional allomorph **k**, which is however much less frequently used.

Returning to the contexts which do not consist of numerals, a few interesting observations can be made. The collocation **ku podivu** behaves as a fixed phrase used as an adverbial of manner meaning *astonishingly enough*, whereas the collocations with the non-vocalised form are used in the literal sense, e.g. **k podivu celého kolektivu lékařů** *to the astonishment of the whole team of doctors*.

The 13 instances of **ku pomoci** *of help, for help* typically co-occur with a verb meaning *be, have, or take*, and are typically not followed by any further subordinated complements, e.g. **NATO nám tu může být ku pomoci svým knowhow** *NATO can be of help for us here with its know-how*, whereas collocations with the non-vocalised form occur more freely with different types of verbs and often take further actants, e.g. **postoj občanů této země k pomoci v umírání** *the standpoint of the citizens in this country towards euthanasia*.

In the instance of **ku Praze** *towards Prague*, the 26 occurrences with the vocalised form are predominantly used in a spatial sense, e.g. **směrem ku Praze** *in the direction towards Prague*, whereas the non-vocalised form almost without exception is preferred in non-spatial senses, e.g. **vztah k Praze** *the relationship to Prague*. However, the non-vocalised form is also found quite a few times in the spatial sense.

Ku příkladu is used as an adverbial phrase meaning *for example*, whereas collocations with the non-vocalised form occur in other contexts, e.g. **připojit k příkladu** *append to the example*.

Ku příležitosti only occurs as an adverbial phrase meaning *on the occasion of*, whereas collocations with the non-vocalised form occur with the very same meaning, as well as in other contexts.

No systematic differences in the usage of **k/ke/ku prospěchu** *to the benefit* have been found. The form **ku** is by far the most common, while **ke** is the most

uncommon. The two instances of **ku vlastnímu** occur in the longer phrase **ku vlastnímu prospěchu**, possibly under the influence of the phrase **ku prospěchu**.

One of the two instances of **ku podpoře** is found in the name of an institution established in 1803, i.e. **Jednota umělců hudebních ku podpoře vdov a sirotků** *The Musicians' Union for the Support of Widows and Orphans*. The other instance seems to be used only as a stylistic variation.

All four instances of **ku lásce** *to love* occur with the verb **zvat** *invite* and the syntactic subject **hlas** *voice*, originating from the beginning of Karel Hynek Mácha's famous poem *Máj* from 1836, i.e. **hrdliččin zval ku lásce hlas** *the turtledove's voice invited to love*. **K lásce** is obviously the normal form in contemporary usage.

In contexts before sibilant followed by consonant, i.e. before **slávě** *glory*, **snížení** *reduction*, **spokojenosti** *satisfaction*, **svému** *one's own*, **škodě** *harm* and **zdraví** *health*, the rate of the form **ke** is, as expected from the earlier findings, well over 90%. Some instances with **k** occur as well, but no pattern can be found. However, most of the very few instances of the form **ku** were found in archaic contexts or in citations from older texts, e.g. **ku slávě boží** *to the glory of God*, found in a text with a religious motivation, or **směřovati ku snížení** *to aim to reduction*, found in a citation from 1922.

To sum up, the form **ku** seems to be used above all to express relative comparison, but it is also frequently used in a few fixed phrases with special meanings before nouns beginning with the consonant clusters /pr/ and /př/. Apart from that, a few more or less random occurrences can be observed, where achieving an archaic effect seems to be the reason for the use of the form **ku**.

Let us now turn to the remaining instances of the vocalised form **ke**, found in contexts with a lower rate of that particular form according to Table 103.

The first context is before words beginning with (č), where the vocalised form **ke** occurs at a rate of 13%. Apart from the four single instances of **ke článku** *to the article*, **ke človíčkům** *to the little people*, **ke člunu** *to the boat*, **ke čtivosti** *to the readability*, higher frequencies are found before the following strings.

ke člen*	70	51%	k člen*	66
ke čten*	75		k čten*	0
ke čtr*	23		k čtr*	0
ke čtv*	71		k čtv*	3
ke čty*	182	92%	k čty*	16

Table 112. Vocalisation of the preposition **k** before words beginning with (č) in the contemporary part of the Czech corpus.

Only the vocalised form is found before words beginning with **čten*** *read*. The

vocalised form also dominates heavily before words beginning with **čtr***, **čtv***, **čty***, all three strings occurring exclusively in words based on the numeral *four*. Finally, both the vocalised and the non-vocalised form are almost equally frequent before words beginning with **člen*** *member*. No distributional pattern could be found for this string.

The next context is before words beginning with ⟨v⟩, where the vocalised form **ke** occurs at a rate of 10%. For this context the number of instances is four times as high as for the preceding context of ⟨č⟩, and the range of different collocational types is enormous, which means that most collocational types are only represented by a couple of instances. Therefore it is impossible to deal with this context in detail, and the data has first been analysed with regard to the two initial graphemes, as shown in the following table.

ke vč*	108	65%	k vč*	59
ke vě*	3	0%	k vě*	1019
ke vh*	2		k vh*	10
ke vch* ⁹⁴	30		k vch*	0
ke vj*	1		k vj*	8
ke vk*	12		k vk*	4
ke vl*	10	1%	k vl*	888
ke vn*	5	5%	k vn*	106
ke vp*	5		k vp*	1
ke vr*	3	0%	k vr*	600
ke vř*	1		k vř*	4
ke vs*	516	93%	k vs*	36
ke vš*	703	87%	k vš*	109
ke vt*	4		k vt*	4
ke vy*	1	0%	k vy*	3,789
ke vz*	570	77%	k vz*	167

Table 113. Vocalisation of the preposition **k** before words beginning with ⟨v⟩ in the contemporary part of the Czech corpus.

The highest number of occurrences, as well as very high relative frequency rates, are found in contexts before ⟨v⟩ followed by a sibilant. On closer investigation of these data, it turns out that the sibilant in the context **vs*** is practically always followed by another consonant, the only exceptions being the lexeme **ves** in the collocational types **ke vsi** *towards the village* and **ke vsím** *towards the villages*. This lexeme does not occur in collocations with the non-

⁹⁴ As ⟨ch⟩ represents the phoneme /x/, a string of three graphemes has been preferred here.

vocalised form of the preposition. The distribution in contexts where ⟨vs⟩ is followed by yet another consonant very much resembles the distribution in the context before ⟨s⟩ followed by a consonant, exhibiting the form **ke** in about 90% of instances, but with no other obvious pattern.

Before ⟨vz⟩ the situation is somewhat different. Here, many of the contexts triggering the vocalised form of the preposition consist of ⟨vz⟩ followed by a vowel. This is so in 68 instances, which gives a vocalisation rate of 43% in this context. The corresponding rate before ⟨vz⟩ followed by a consonant is 87%, which is close to the corresponding rate before ⟨z⟩ followed by a consonant.

Finally, before ⟨vš⟩, one single stem dominates the data, namely /vš/ meaning *all, whole*. Out of a total of 703 instances, 652 collocations contain the three pronominal forms **všem**, **všemu** and **vší**. Apart from that, a few instances occur with different adjectives and nouns, all beginning with the morpheme /vš/, e.g. the single instances **ke všelidovému** *to all mankind's*, **ke všestrannosti** *to the versatility* and the quite frequent **ke všeobecn*** (43x) *to the general*. For comparison, only **všem**, of these three pronominal forms, occurs once with the non-vocalised form of the preposition, and the remaining 108 instances involve other lexemes, whereof the absolute majority however begin with the root {vš} *all*.

In the next context, which is ⟨v⟩ followed by the affricative ⟨č⟩, with many occurrences of the vocalised form **ke**, the data are once again dominated by a single root, namely {včer} *yesterday*. 104 out of the 108 occurrences with the vocalised form **ke** before ⟨vč⟩ involve lexemes with this root. As a comparison, 48 out of the 59 occurrences with the non-vocalised form **k** also involve the same root. This means that the vocalised form is clearly preferred before the root {včer}, whereas the non-vocalised form is clearly preferred elsewhere.

Again for the context of ⟨vch⟩, only one lexeme is involved, namely **vchod** *entrance*, and its derived adjective **vchodový**. All 30 occurrences of **ke** in this context are found before this lexeme and its derived adjective. No occurrences are present in the corpus of the non-vocalised form of the preposition before words beginning with ⟨vch⟩.

For the remaining contexts beginning with ⟨v⟩ the data are very scanty and no further conclusions can be made. Two interesting observations should be pointed out, however. Firstly, the vocalised form of the preposition practically does not occur before a single consonant followed by a vowel. The only actual exception is a single occurrence before ⟨vy⟩, whereas the context ⟨vě⟩ actually corresponds to the phonological string /vje/. Secondly, the quite high frequency of the vocalised form **ke** before ⟨vk⟩, ⟨vp⟩ and ⟨vt⟩ may possibly be explained by the plosive character of the second element, although no other pattern can be observed with regard to the lexemes involved.

Next, the context before ⟨h⟩ will be investigated.

ke hl*	21	3%	k hl*	750
ke hm*	1		k hm*	10
ke hn*	6		k hn*	58
ke hr*	82	10%	k hr*	753
ke hř*	136	90%	k hř*	15
ke hv*	34		k hv*	25

Table 114. Vocalisation of the preposition **k** before words beginning with (h) in the contemporary part of the Czech corpus.

The data concerning contexts with a low frequency of vocalisation of the preposition do not reveal any interesting patterns apart from the observation that the only collocational type in which the rate of the vocalised form **ke** is higher than the non-vocalised **k** is **ke hrám** (11/8) *to the games*.

In the context with the highest frequency of the vocalised form of the preposition, i.e. before ⟨hř⟩, the situation is absolutely dominated by two collocational types, i.e. **ke hře** (114/0) *to the game* and **ke hřbitovu** (11/3) *towards the cemetery*. There is also one instance of the derived adjective **ke hřbitovní** (1/0) *towards the cemetery's*. One collocation is equally frequent with both forms of the preposition, namely **ke hřišti** (5/5) *towards the playground*, whereas the rest of the occurrences with the vocalised form are insignificant.

It should be noticed that once again, as with /vš/ *all, whole* above, roots exhibiting the allomorphs /hr/ and /hř/ lacks a vocalic element. Further, the lexeme **hřbitov** begins with the extremely uncommon consonant cluster /hřb/.

The next frequent context of vocalisation to be investigated is ⟨d⟩.

ke dějinám	1		k dějinám	31
ke dlouh*	7	3%	k dlouh*	200
ke dluž*	1		k dluž*	11
ke dne*	10	3%	k dne*	301
ke dni	188	100%	k dni	0
ke dnu	213	100%	k dnu	0
ke dnům	1		k dnům	0
ke dra*	4	2%	k dra*	167
ke dráze	2		k dráze	7
ke drogám	2		k drogám	47
ke druhé	68	49%	k druhé	70
ke druhému	106	41%	k druhému	154
ke druhým	5		k druhým	33
ke dřevu	3		k dřevu	0
ke dřívě	3		k dřívě	8
ke dva*	42	38%	k dva*	69
ke dvěma	290	93%	k dvěma	21

ke dveřím	80	k dveřím	0
ke dvěřtě	1	k dvěřtě	1
ke dvřřkům	2	k dvřřkům	0
ke dvoj*	8	k dvoj*	34
ke dvorku	1	k dvorku	0
ke dvoru	4	k dvoru	0
ke dvořák*	2	k dvořák*	8
ke dvou*	22	k dvou*	54
ke džbánu	1	k džbánu	0

Table 115. Vocalisation of the preposition **k** before words beginning with ⟨d⟩ in the contemporary part of the Czech corpus.

Here, the most prominent collocational types are **ke dni** *to the day*, **ke dnu** *to the bottom*, **ke dveřím** *to the door*, where only the vocalised form of the preposition is attested in the corpus. The first two collocational types once again contain nouns having a root without a vowel, i.e. {dn} (masculine) *day* and {dn} (neutre) *bottom*. In the case of **ke dveřím** no phonological reason for the vocalised form can be found, but it is worth noting that this lexeme has already been encountered with the vocalised form of many prepositions in Slovak and Polish as well. There are also a few other collocational types that only occur with the vocalised form of the preposition, but the numbers of occurrences are too small to be significant, i.e. **ke dnům** (1/0) *to the days*, **ke dřevu** (3/0) *to the tree*, **ke dvřřkům** (2/0) *to the little door*, **ke dvorku** (1/0) *to the little yard*, **ke dvoru** (4/0) *to the yard*, **ke džbánu** (1/0) *to the jug*.

In the case of **ke dvěřma** (290/21) *to the two*, the vocalised form of the preposition is clearly dominant.

In the collocational types **ke druhě** (68/70), **ke druhému** (106/154), both meaning *to the other*, *to the second*, as well as before words beginning with **dva*** (42/69) *two*, the vocalised form of the preposition is almost as common as the non-vocalised. The same is likely to be true of **ke dvěřtě** (1/1) *two hundred*, even though the data are extremely scanty. In other contexts with the root {dv} *two*, there seems to be a lesser tendency towards the vocalised form, i.e. **ke dvoj*** (8/34) *to the double* and **ke dvou*** (22/54) *to the two*.

For the remaining occurrences of the vocalised form before ⟨d⟩, the data are scanty and it is impossible to tell if these occurrences are evidence of any patterns.

The next context to be investigated is before ⟨m⟩.

ke mě*	7	3%	k mě*	260
ke mě	6		k mě	0
ke ml*	1	1%	k ml*	120
ke mletí	1		k mletí	2

ke mn*	370	64%	k mn*	212
ke mně	369	100%	k mně	0
ke mřížím	1		k mříži	1
ke mstě	1		k mstě	0
ke mši	4		k mši	0
ke mzdám	15		k mzdám	1
ke mzdě	19		k mzdě	0
ke mzdov*	4		k mzdov*	8

Table 116. Vocalisation of the preposition *k* before words beginning with ⟨m⟩ in the contemporary part of the Czech corpus.

As could be expected, only the vocalised form of the preposition occurs in collocations with the 1st person singular personal pronoun, i.e. **ke mě/ke mně**⁹⁵ *to me, towards me*.

The collocations **ke mstě** *to the revenge*, **ke mši** *to mass*, **ke mzdám** (pl.) *to the wages*, **ke mzdě** (sg.) *to the wage*, and **ke mzdov*** (adj.) *to the wages'*, have in common that they involve roots which do not contain any vowel, i.e. {mst}, {mš} and {mzd}. In the above data, the vocalised form of the preposition is absolutely predominant before root nouns where the stem of the word lacks a vowel, whereas the non-vocalised form is preferred before the derived adjective, i.e. when the stem of the word contains a vowel in the derivational suffix.

The single instances **ke městu** (1/66) *to the town*, **ke mletí** (1/2) *to grinding*, **ke mnohočetné** (1/0) *to the manifold*, **ke mřížím** (1/0) *to the grating*, might be based on individual choices. In the corpus, there are many other collocations with the non-vocalised form of the preposition in the same phonological context, as well as before other words based on the same root, e.g. 194 instances of **k mnoh*/k množ*** *to many*.

The next context to be investigated is before ⟨c⟩. For this context, it immediately becomes clear that the majority of instances with the vocalised form of the preposition occur before words of foreign origin in which the initial ⟨c⟩ is pronounced [k], e.g. **ke Camile**, which is absolutely in accordance with the already established findings. The following table therefore only comprises data where the initial ⟨c⟩ or ⟨ch⟩ is pronounced as [ts] or [x].

ke cti*	102	98%	k cti*	2
ke cti	99		k cti	0

⁹⁵ **Ke mě** is a spelling mistake due to the fact that **mě** (acc., gen.) and **mně** (dat., loc.) are only orthographic variants signifying different cases. The pronunciation is the same in both instances.

ke ctnostem	1	k ctnostné	1
ke cvič* [*]	16	k cvič* [*]	1
ke cvičení	9	k cvičení	0
ke cvičením	5	k cvičením	0
ke chl* [*]	6	k chl* [*]	32
ke chm* [*]	1	k chm* [*]	1
ke chr* [*]	5	k chr* [*]	22
ke chv* [*]	10	k chv* [*]	5
ke chvále	6	k chvále	1

Table 117. Vocalisation of the preposition **k** before words beginning with ⟨ch⟩ in the contemporary part of the Czech corpus.

In these collocational types the relative frequency of the vocalised form of the preposition is generally quite high, which may partly be caused by the complex nature of the affricate phoneme /c/, as well as the closeness in articulatory place between the phoneme /x/ and the /k/ of the preposition. The quite large variation before /x/ is however noteworthy.

The vocalised form of the preposition seems to be obligatory in the collocational type **ke cti** *to the honour*, as well as before the lexeme **cvičení** *exercise*. It also seems that the vocalised form is highly preferred in the collocational type **ke chvále** (6/1) *to the praise*.

In the remaining collocational types several different lexemes occur just once or a few times each. No patterns were noticed.

The next context to be investigated is before words beginning with ⟨b⟩.

ke bř* [*]	52	k bř* [*]	44
ke břehu	36	k břehu	1
ke břehům	11	k břehům	23
ke Březině	4	k Březině	0
ke Břvům	1	k Břvům	0

Table 118. Vocalisation of the preposition **k** before words beginning with ⟨b⟩ in the contemporary part of the Czech corpus.

In this context only four different word forms beginning with ⟨bř⟩ trigger the use of the vocalised form of the preposition. It is evident that the vocalised form is highly preferred before the singular noun in **ke břehu** (36/1) *towards the shore*, but less so before the plural form **ke břehům** (11/23) *towards the shores*. With the two forms of proper names **Březině** and **Břvům**, only the vocalised form of the preposition is attested in the corpus, but the data are too scanty to draw any conclusions.

The next context to be investigated is before words beginning with ⟨t⟩.

ke tch*	1		k tch*	2
ke tk*	1		k tk*	3
ke tm*	2		k tm*	9
ke tr*	7	0%	k tr*	1,742
ke tř*	346	72%	k tř*	136
ke třem	179	94%	ke třem	11
ke třetí	33		ke třetí	8
ke třetím	2		ke třetím	1
ke třetímu	42		ke třetímu	22
ke tři*	90	52%	ke tři*	84
ke tu*	1	1%	k tu*	178
ke tv*	14	4%	k tv*	323
ke tvorbě	6		k tvorbě	87

Table 119. Vocalisation of the preposition **k** before words beginning with ⟨t⟩ in the contemporary part of the Czech corpus.

Here the most frequent collocational types with the vocalised form of the preposition all involve different forms of the numeral *three*. For the collocational type with the basic cardinal numeral **ke třem** (179/11) the vocalised form of the preposition clearly dominates. Before the ordinal numeral **třetí*** *third*, the vocalised form of the preposition is somewhat more common than the non-vocalised, whereas before other words beginning with **tři*** both forms of the preposition are almost equally common. For the remaining collocations with **ke** before ⟨t⟩, no specific patterns could be observed.

The next context to be investigated is before words beginning with ⟨j⟩.

ke jm*	69	k jm*	9
ke jmění	1	k jmění	0
ke jmenování	32	k jmenování	7
ke jměnu	24	k jměnu	0
ke jměním	9	k jměním	0

Table 120. Vocalisation of the preposition **k** before words beginning with ⟨j⟩ in the contemporary part of the Czech corpus.

The only consonant cluster involved is ⟨jm⟩, before which the vocalised form of the preposition is highly preferred, especially before the noun **jměno** *name*, both in the singular and in the plural.

The next context to be investigated is before words beginning with ⟨l⟩.

ke lh*	9	k lh*	0
ke lhaní	4		

ke lhůtám	3		
ke lsti	3	k ls*	0
ke lž*	17	k lž*	2
ke lži	13	k lži	0
ke lžím	2	k lžím	1
ke lživé	2	k lživé	1

Table 121. Vocalisation of the preposition **k** before words beginning with ⟨l⟩ in the contemporary part of the Czech corpus.

Also before the consonant clusters ⟨lh⟩, ⟨ls⟩ and ⟨lž⟩, the vocalised form of the preposition is very highly preferred. It should be noted that two more root nouns without a vowel in the stem occur in the above table, i.e. ⟨lst⟩ *trick* and ⟨lž⟩ *lie*.

The next context to be investigated is before ⟨r⟩.

ke rv*	16	k rv*	1
ke rvačce	14	k rvačce	1

Table 122. Vocalisation of the preposition **k** before words beginning with ⟨r⟩ in the contemporary part of the Czech corpus.

The only consonant cluster involved is ⟨rv⟩ and the vocalised form of the preposition is clearly preferred.

The next context to be investigated is before ⟨p⟩.

ke pl*	1	0%	k pl*	777
ke pr*	4	0%	k pr*	7,082
ke prospěchu	4		k prospěchu	25
ke př*	5	0%	k př*	5,223
ke ps*	8	8%	k ps*	98
ke psaní	5		k psaní	25
ke psovi	1		k psovi	1
ke psům	2		k psům	0
ke pštrosům	2		k pš*	1

Table 123. Vocalisation of the preposition **k** before words beginning with ⟨p⟩ in the contemporary part of the Czech corpus.

The only interesting collocations within this context are the root noun ⟨ps⟩ *dog*, without any vowel, and the extremely complex initial consonant cluster in **ke pštrosům** *to the ostriches*. For the remainder of these data no special patterns could be observed.

For the remaining two contexts, i.e. before ⟨f⟩ and ⟨n⟩, the data **ke fiasku** (1/3) *to the fiasco*, **ke Flekům** (1/0) *to Fleků*, **ke Francii** (1/28) *to France*, **ke**

několika (1/163) *to some*, **ke nížinným** (1/0) *to the lowland's*, no interesting observations could be made. It should however be pointed out that **k fl*** occurs 25 times and **k ní*** 749 times.

Summing up, the many individual collocations with the vocalised form of the preposition give an indication of a somewhat different situation in Czech than in Slovak and Polish. It seems that in certain contexts Czech offers a certain amount of freedom for individual stylistic preferences when it comes to the choice between the vocalised and the non-vocalised forms of the preposition **k/ke/ku**.

8.1.2.2. **v(e)**

All instances of the vocalised form of the preposition **v(e)** were extracted from the corpus and sorted with respect to the first grapheme of the following word. Then all occurrences of the non-vocalised form **v** occurring in the same contexts were counted for comparison.

During this process, quite a lot of instances were encountered that did not represent instances of the preposition, such as wrongly inserted spaces, e.g. **dří ve** instead of **dříve** *earlier*, the abbreviation **VE** or **ve** with different meanings, instances where **ve** was uninterpretable, e.g. **ve arény ryku**, instances in English, e.g. **I've**, instances with a misprinted diacritical sign, e.g. **ve'cer** instead of **večer** *evening*, as well as instances due to misprints in Slovak words, e.g. **ve?ké** instead of **velké** *big*. Further, the preposition **ve** frequently occurred before the Roman numerals I, II, III, IV and XIII, as well as before html code, e.g. **ve (b)**, wrongly before another preposition, e.g. **ve nad Černými lesy** *in above Černé lesy*, or before incorrectly spelled words, e.g. **ve redověku** instead of the correct **ve středověku** *in the middle ages*. All such occurrences of the form **ve** have been omitted from the data in the following table.

For the non-vocalised form of the preposition, there were too many abbreviations and divided words to eliminate them all manually. They have therefore only been excluded from the data in the contexts where the number of non-vocalised occurrences was lower than the number of vocalised occurrences in the same context, e.g. before abbreviations such as **v FBI** *in FBI* which would most probably be pronounced with an initial [ef], due to spelling mistakes, e.g. **v vozduší** instead of the correct **v ovzduší** *in the atmosphere* or **v vozů** instead of the correct **z vozů** *from the carts*, as **v** cannot possibly be followed by the noun **vozů** because it does not have the appropriate case ending, or when **v** itself was not an instance of the preposition, but an abbreviation.

All the above mentioned instances have been disregarded in the following table.

ve a*	2	0%	v a*	25,139
ve b*	31	0%	v b*	54,174
ve c*	2,513	9%	v c*	25,370
ve č*	9,601	19%	v č*	40,499
ve d* /ve d**	12,780	17%	v d* / v d**	64,092
ve e*	2	0%	v e*	13,906
ve f*	1,823	98%	v f*	30
ve g*	5	0%	v g*	4,690
ve h*	1,889	5%	v h*	35,290
ve j*	737	2%	v j*	37,841
ve k*	6,539	10%	v k*	62,259
ve l*	183	0%	v l*	43,973
ve m*	5,015	9%	v m*	53,129
ve n* /ve ň*	29	0%	v n* / v ň*	112,658
ve o*	6	0%	v o*	55,350
ve p*	2,463	1%	v p*	224,473
ve q*	1	1%	v q*	76
ve r*	160	0%	v r*	86,731
ve ř*	6	0%	v ř*	8,809
ve s*	91,275	57%	v s*	68,114
ve š*	9,355	63%	v š*	5,401
ve t* /ve t**	5,667	6%	v t* / v t**	92,582
ve u* /ve ú*	11	0%	v u* / v ú*	39,211
ve v*	67,696	100%	v v*	61
ve w*	2,343	100%	v w*	3
ve z*	10,082	18%	v z*	46,058
ve ž*	271	3%	v ž*	8,532
ve 0*	84	12%	v 0*	640
ve 1*	2,258	24%	v 1*	7,328
ve 2*	3,281	87%	v 2*	509
ve 3*	1,220	78%	v 3*	340
ve 4*	1,052	91%	v 4*	109
ve 5*	2	0%	v 5*	1,343
ve 6*	2	0%	v 6*	1,209
ve 7*	3	0%	v 7*	1,196
ve 8*	1	0%	v 8*	982
ve 9*	1	0%	v 9*	697

Table 124. Vocalisation of the preposition **v** in the Czech corpus.

Already, based on the above table and the observations previously made about the Czech preposition **k(e)/ku**, as well as about the Slovak and Polish

prepositions, a few quite interesting observations can be made.

Vocalisation of the preposition **v(e)** seems to be obligatory before the phonemes /v/ and /f/. The rate of vocalisation before the sibilants ⟨s⟩ and ⟨š⟩ is also quite high, but before ⟨z⟩ and ⟨ž⟩ it is unexpectedly low. The digits **2, 3, 4** trigger a very high frequency of vocalisation.

The digit **1** triggers a medium frequency of vocalisation, but this might be due to the fact that this digit occurs in combinations which, when pronounced, begin with quite different morphemes, i.e. **1, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 100, 1000**. The combinations **12 (dvanáct), 13 (třináct), 14 (čtrnáct)** could be expected to behave similarly to **2 (dva/dvě), 3 (tři), 4 (čtyři)**, whereas **100 (sto)** begins with a sibilant-initial consonant cluster. Finally, it is interesting that the digit **0** also triggers a certain amount of vocalisation of the preposition **v(e)**, unlike in the case of **k(e)/ku**. This phenomenon turns out to be due to time expressions of the type **ve 03.05** (i.e. **ve tři nula pět**) at 3.05, where 2, 3 or 4 occur as the second digit, which in most cases can be assumed to be the first digit to be read out.

We will now examine these contexts more closely, and we begin with the instances which do not comply with the seemingly obligatory principle of vocalisation before /f/ and /v/. 16 of these 30 instances of vocalisation occur before proper names, mainly geographical, most of which are of foreign origin, e.g. **v Fieberbrunnu** in *Fieberbrunn*. The relative rate is however so low that these few occurrences must be interpreted as writing mistakes or instances of a deliberate personal choice.

When it comes to lack of vocalisation before ⟨v⟩, the picture is somewhat different. Here, 38 out of 61 occurrences were found in the older part of the corpus, whereas only two out of 30 occurred before ⟨f⟩. This implies that collocations with the preposition **v** before ⟨v⟩ might have been slightly more tolerable a century or more ago. All the collocations are represented by one single occurrence, except for **v večerní době** in *the evening time* and **v věcech** in *the things*, which both occur twice. Finally, all three instances of the non-vocalised form of the preposition before ⟨w⟩ are found in the context of foreign geographical names.

We will now look more closely into the situation before words beginning with a sibilant. All the occurrences of the preposition **v(e)** are divided into contexts where the following words begin with a sibilant followed by a consonant (C) or by a vowel (V). As abbreviations are normally read with the insertion of vowels, they have been treated as instances of a sibilant followed by vowel. Words beginning with a sibilant followed by an /r/ or an /l/, which constitutes the syllable nucleus in the same way as a vowel, have been treated as a sibilant followed by a vowel. It is very interesting to note that in this type of context, i.e. sibilant followed by syllabic /r/ or /l/, no instances of vocalisation of the preceding preposition have been found in the corpus. Finally, words beginning with English ⟨sh⟩ for [ʃ], German ⟨sch⟩ for [ʃ] and

Hungarian ⟨sz⟩ for [s] directly followed by a vowel have also been treated as a sibilant followed by a vowel.

ve sC*	91,263	98%	v sC*	2,191
ve šC*	9,352	98%	v šC*	160
ve zC*	10,056	96%	v zC*	387
ve žC*	265	91%	v žC*	25
ve sV*	12	0%	v sV*	65,923
ve šV*	3	0%	v šV*	5,241
ve zV*	26	0%	v zV*	45,671
ve žV*	6	0%	v žV*	8,507

Table 125. Vocalisation of the preposition *v* before sibilants.

It thus becomes clear that the differences in frequency before the four sibilants are not due to the sibilants themselves, but rather to the frequency with which they occur word-initially followed by a consonant or a vowel. It turns out that the preposition *v(e)* is vocalised before all four sibilants at a very high rate if the sibilant is followed by a consonant, and that it is practically never vocalised if the sibilant is followed by a vowel or a vocalic /r/ or /l/.

Many of the contexts in which the preposition is not vocalised when followed by a sibilant-initial consonant cluster, are found in the older parts of the corpus. In order to investigate possible changes in the distribution over time, the above data are divided into occurrences in the contemporary and the older parts of the corpus. The results are displayed in the following two tables.

ve SC*	104,260	99%	v SC*	668
ve SV*	43	0%	v SV*	118,619

Table 126. Vocalisation of the preposition *v* before sibilants in the contemporary part of the Czech corpus.

ve SC*	6,513	76%	v SC*	2,095
ve SV*	4	0%	v SV*	6,723

Table 127. Vocalisation of the preposition *v* before sibilants in the older part of the Czech corpus.

The subdivision of the data into the two tables above clearly shows that the diachronic aspect is very important in these types of context. In contemporary usage the obvious tendency is always to vocalise before a sibilant-initial consonant cluster, but not to vocalise before an initial sibilant followed by a vowel or an /r/ or /l/ which constitutes the nucleus of a syllable. A century or more ago, the situation before sibilant consonant clusters was not so stable, and both the vocalised (76%) and the non-vocalised (24%) forms of the

preposition were in frequent use.

We will now go on to investigate the remaining contexts which exhibit a significant rate of vocalisation of the preposition **v(e)**. First, the data are revised in order to limit them to the contemporary part of the corpus, and all collocations in which the preposition is followed by an abbreviation are disregarded, as it is impossible to know with certainty how the abbreviation was intended to be read. The collocation **ve USA** *in the USA* may serve as an example. It is very unexpected to find the vocalised form of the preposition before a vowel, and the most plausible reason for the vocalisation in this collocation is that it is intended to be read as **ve Spojených státech** *in the United States*.

The revised data, excluding contexts before (k), (g) or a sibilant, based only on the contemporary part of the Czech corpus, are displayed in the following table.

ve č*	9,370	20%	v č*	38,573
ve d*	12,017	17%	v d*	59,344
ve c*	2,308	9%	v c*	23,559
ve k*	5,887	9%	v k*	56,718
ve m*	4,370	8%	v m*	49,521
ve t*	5,239	6%	v t*	82,231
ve h*	1,748	5%	v h*	32,310
ve j*	473	1%	v j*	34,522
ve p*	2,329	1%	v p*	213,093
ve q*	1	1%	v q*	73
ve a*	0	0%	v a*	20,091
ve b*	3	0%	v b*	51,436
ve ḍ*	0	0%	v ḍ*	151
ve g*	5	0%	v g*	4,480
ve l*	157	0%	v l*	40,994
ve n*	13	0%	v n*	102,363
ve ň*	0	0%	v ň*	0
ve o*	3	0%	v o*	52,427
ve r*	141	0%	v r*	83,454
ve ř*	1	0%	v ř*	8,326
ve ṭ*	0	0%	v ṭ*	6
ve u*	0	0%	v u*	15,800
ve ú*	0	0%	v ú*	21,506

Table 128. Vocalisation of the preposition **v** in selected contexts in the contemporary part of the Czech corpus.

These data will now be examined in more detail, beginning with the context

that exhibits the highest frequency of vocalisation of the preceding preposition, namely the context before ⟨č⟩, where the rate of the vocalised form is 20%.

ve če*	1	0%	v če*	27,972
ve čk*	5		v čk*	0
ve čl*	47	6%	v čl*	748
ve čp*	1		v čp*	0
ve čt*	9,302	99%	v čt*	52

Table 129. Vocalisation of the preposition *v* before ⟨č⟩ in the contemporary part of the Czech corpus.

It seems clear that consonant clusters once again constitute the decisive factor. Therefore, the rest of the collocations containing the non-vocalised form of the preposition were also checked for occurrences of the form *k* before other consonant clusters, but no such occurrences were encountered in the corpus. A closer look at the data also reveals that the 47 instances of *ve čl** as well as the 52 instances of *v čt** contain a large variety of collocational types. The data therefore evidently suggest that the choice of form of the preposition is not bound to certain lexemes, but rather triggered by the character of the initial consonant cluster of the following word.

The next context to be investigated is before ⟨d⟩, where the overall rate of vocalisation is 17%.

ve da*	2	0%	v da*	6,420
ve dl*	7	1%	v dl*	1,272
ve dm*	1		v dm*	2
ve dn*	1,042	41%	v dn*	1,493
ve dr*	6,675	55%	v dr*	5,364
ve dř*	25	12%	v dř*	187
ve dv*	4,212	92%	v dv*	365
ve dz*	1		v dz*	0
ve dž*	3	2%	v dž*	171

Table 130. Vocalisation of the preposition *v* before ⟨d⟩ in the contemporary part of the Czech corpus.

The most interesting are the many instances and the high rate of vocalisation before ⟨dv⟩, ⟨dr⟩, and ⟨dn⟩. It turns out, however, that the phonological context does not seem to be the major triggering factor. Instead, the preferences for different collocations seem to be dependent on which lexeme or root is involved.

The simplest case is found before ⟨dn⟩. Here vocalisation is preferred to a

very high extent before the nouns **den** *day* and **dno** *bottom*, which both lack a vowel in their homonymous stem {dn}: **ve dne** (303/0) *in day time*, **ve dně** (6/0) *in the bottom*, **ve dnech** (709/3) *in the days*, **ve dni** (3/0) *in the day*, **ve dnu** (9/1) *in the day*, **ve dny** (6/2) *in the days*. Quite contrarily, the non-vocalised form is highly preferred before other stems based on the same root: **v dnes** (49/3) *in today*, **v dnešní*** (1426/2) *in today's*. Besides these collocations, there are actually only nine other instances with the non-vocalised form of the preposition before {dn}, namely **v dnešku** (3/0) *in today's day*, **v Dněpru** (1/0) *in the Dnieper*, **v Dněpropetrovsku** (3/0) *in Dnipropetrovsk* and **v Dněpropetrovské** (2/0) *in Dnipropetrovsk's*.

The situation before {dv} is somewhat more complex. Only the vocalised form is found before stems meaning *door*, i.e. **ve dveřích** (170/0) *in the door*, **ve dveřních** (5/0) *in the door's*, **ve dvířkách** (2/0) *in the little door*. No other collocations with the preposition **v(e)** and this root occur in the corpus. High frequencies are also found before different forms of the cardinal numeral root {dv} *two*, i.e. **ve dva*** (872/77) *in two*, **ve dvě*** (226/2) *in two*, **ve dvoj*** (232/89) *in double* and **ve dvou*** (2217/148) *in two*, as well as before the root {dvor} *courtyard*, namely **ve dvoře** (388/0) *in the courtyard*, **ve dvoraně** (12/0) *in the hall*, **ve Dvorcích** (4/5) *in Dvorky*, **ve dvorech** (11/0) *in the courtyards*, **ve dvorku** (1/0) *in the little yard*, **ve dvorní*** (3/1) *in the court's*, **ve dvorských** (1/2) *in the courtly*, **ve Dvořákov*** (61/32) *in Dvořák's*, **ve Dvořisku** (4/2) *in Dvořisko*, **ve Dvořišti** (3/1) *in Dvořiště*, whereas only the non-vocalised form occurs in the collocational types **v dvoranu** (1/0) *into the hall*, **v dvorci** (2/0) *in the residence*, **v dvorné** (1/0) *in the courteous*, and **v dvořištské** (1/0) *in Dvořiště's*. It seems that the vocalised form of the preposition dominates quite heavily before the two roots {dv} and {dvor}, and that the frequency of the non-vocalised form may be somewhat higher with derivations, especially the less frequent ones, whereas in the collocations **ve dvoře**, **ve dvorech** and **ve dvorku**, involving the basic root noun, only the vocalised form is encountered. It should also be noted that apart from the collocations mentioned, there are actually no other instances in the corpus with the non-vocalised form of the preposition before {dv}.

The situation before {dr} also seems to be fairly straightforward. The most frequent collocational types contain the three locative case forms of the adjective **druhý** *second, other*, i.e. **ve druhé** (2990/1141), **ve druhém** (3634/854), and **ve druhých** (19/12). The non-vocalised form of the preposition prevails, however, before other forms of the same lexeme as well as before other lexemes with the same root, i.e. **ve druhého** (1/2), **ve druhou** (2/12), **ve druhý** (8/15), as well as compositions containing **ve druho*** (13/100). Apart from these collocational types, a few individual occurrences were found, namely **ve druháku** (1/0) *in second grade*, **ve dramaturgickém** (1/5) *in the dramaturgical*, **ve dresu** (1/968) *in the sports outfit*, **ve družině** (1/18) *in the after-school care centre*, **ve družstevním** (1/27) *in the cooperative's*, **ve družstvech** (2/57) *in the cooperative*. There are however many other collocational types with only the

non-vocalised form of the preposition, e.g. **v drobných** *in small*, **v drogerii** *at the chemist's*, which apparently represent the unmarked usage.

The collocations in contexts before ⟨dř⟩ do not reveal any interesting patterns apart from the collocations **ve dřevě** (12/0) and **ve dřevu** (1/0), both involving different forms of the locative case of the noun **dřevo** *wood*. For other collocations with derivations from the same root, the data are very scanty, namely **ve dřevařském** (1/2) *in the timber*, **ve Dřevíčském** (1/0) *in Dřevíč's*, **ve dřevotříšce** (1/0) *in the chipboard*. The vocalised form also occurs in a few other collocations, **ve dřepu** (1/9) *with bent knees*, **ve dřezu** (4/6) *in the sink*, **ve dřív/dřive** (4/12) *in the earlier*, **ve dřívější** (1/14) *in the previous*, but the frequency is quite low. The non-vocalised form occurs in a number of other collocational types, and it seems that the non-vocalised form is the unmarked choice before ⟨dř⟩.

The three collocations with the vocalised form of the preposition in the context of ⟨dž⟩ are all instances where these two initial consonants are followed by yet another consonant, namely **ve džbáně** (1/0) *in the jug*, **ve džbáněch** (1/0) *in the jugs*, and **ve Džmurův** (1/0) *into Džmur's*, whereas in the 171 collocations with the non-vocalised form of the preposition, the initial ⟨dž⟩ is followed by a vowel. A reasonable interpretation of these data is that /dž/ constitutes a single phoneme, not a consonant cluster, hence there is no need for vocalisation before this single initial consonant followed by a vowel.

The collocations in the context of ⟨dm⟩, **ve Dmitrově** (1/1) *in Dmitrov*, **v Dmytrykově** (1/0) *in Dmytrykov* and ⟨dz⟩, **ve Dzbelu** (1/0) *in Dzbel* are so few that the high degree of vocalisation may not necessarily be significant.

The few vocalisations in the contexts of ⟨da⟩ and ⟨dl⟩ exhibit no patterns. They may readily be interpreted as writing mistakes or deliberate individual stylistic choices.

We will now continue to investigate collocations with the vocalised form of the preposition before words beginning with ⟨k⟩.

ve kb*	93	85%	v kb*	16
ve kd*	12		v kd*	38
ve kl*	1	0%	v kl*	6,012
ve km*	1		v km*	23
ve kn*	1	0%	v kn*	1,145
ve kr*	2	0%	v kr*	7,533
ve kř*	32	4%	v kř*	802
ve kš*	3		v kš*	2
ve kt*	5,730	92%	v kt*	481
ve kv*	11	0%	v kv*	3,890

Table 131. Vocalisation of the preposition **v** before ⟨k⟩ in the contemporary part of the Czech corpus.

First, it may be observed that the higher frequencies of vocalisation are found before consonant clusters where the second element is a plosive, namely /t/, /b/, /d/. Lower frequencies are found when the second element is a fricative, whereas the lowest frequencies are found when the second element is a sonorant.

We will begin by investigating the context of ⟨kt⟩, which exhibits the highest number of vocalisations. Only two roots occur in all these collocations, namely the pronominal root {kter} *which (one)* and a placename, **Ktoře**: **ve kter*** (5,729/480) and **ve Ktoře** (1/1) in *Ktora* (Israel) respectively. A closer investigation of the collocations containing the pronominal root suggests that when the pronominal stem occurs in different case forms of the basic, underived lexeme **který** *which*, the vocalised form of the preposition dominates to an even higher degree, specifically 95% (5,723/300), whereas in collocations where different derivational suffixes are added to the basic pronoun, the non-vocalised form of the preposition is preferred to a very high degree, namely **v kter*koli(v)** (174/6) in *whatever*, **v kter*si** (1/0) in *a certain*, **v kter*ž(to)** (5/0) in *which*.

The next context of interest is before ⟨kb⟩. Only in the collocational type **ve Kbelích** (80/4) in *Kbely* is the vocalised form clearly preferred, whereas in the remaining collocational types, both forms of the preposition are more or less equally frequent, namely **ve kbelících** (4/5) in *the buckets*, **ve kbelíku** (7/5) in *the bucket*, **ve kbelské** (1/0) in *Kbely's* and **ve Kbelu** (1/2) in *Kbel*.

Also for collocations in the context of words beginning with ⟨kd⟩, only one collocational type shows a high degree of preference for the vocalised form of the preposition, namely **ve Kdyni** (11/1) in *Kdyně*. In other collocations almost no instances of vocalisation are attested: **ve kdysi** (1/21) in *the once*, **v kdejak*** (11/0) in *several*, **v kdo** (2/0) in *who*, **v kdovíjak*** (1/0) in *who knows which*.

The two remaining contexts with more than a very few instances of vocalisation are ⟨kř⟩ and ⟨kv⟩. The vocalisations before ⟨kř⟩ are dominated by **ve křoví** (21/43) in *the bushes*, where the non-vocalised form of the preposition is dominant. However, in the other two collocational types exhibiting instances with the vocalised form of the preposition, namely **ve Křtěnově** (4/0) in *Křtěnov* and **ve Křtinách** (7/0) in *Křtiny*, only the vocalised form of the preposition is attested in the corpus, but in a large number of other collocational types involving the context of ⟨kř⟩, there are a total of 802 instances with the non-vocalised form of the preposition.

The 11 instances of the vocalised form of the preposition before ⟨kv⟩ are spread over 10 different collocational types. The low frequency and the distribution of these instances, as well as the insignificant number of occurrences before the remaining consonant clusters beginning with a ⟨k⟩ indicates that there are no further tangible distributional patterns to be found before ⟨k⟩. The only findings among these scanty data worth mentioning are

the three instances with the vocalised form of the preposition before the lexeme **kšeft** *business*: **ve kšeft** (1/0) *into business*, **ve kšeftě** (1/0) *in the business*, **ve kšeftu** (1/0) *in the business*. In the same context, only two occurrences of the non-vocalised form were encountered: **v kšandách** (1/0) *in braces*, **v kšiltovce** (1/0) *in a peaked cap*. This indicates a possible preference for the vocalised form before the lexeme **kšeft**.

The next orthographical context (c) encompasses two phonological contexts, namely [ts] and [x]. The relevant data for each of these contexts are 93 instances of **ve [ts]*** versus 18,969 **v [ts]***, i.e. a vocalisation rate of 0.5%, and 2,217 instances of **ve [x]*** versus 4,682 instances of **v [x]***, i.e. a rate of 32%.

After removing the [x] contexts, as well as a number of spelling mistakes, collocations involving abbreviations and words of foreign origin beginning with ⟨cl⟩, ⟨cr⟩, and ⟨cz⟩ and pronounced [kl], [kr], and [ʧ], all the remaining instances of the vocalised form of the preposition before [ts], as well as all the instances of the non-vocalised form of the preposition in the same context, are presented in the following table.

ve celé	1	v celé*	1,805
ve cloně	2		0
ve Ctěnicích	9		0
ve Ctěnickém	1		0
ve Ctesifonu	1		0
ve ctíhodné	1		0
ve ctíhodném	1		0
ve ctiteli	1		0
ve ctnostech	1		0
ve cvičebně	1		0
ve cvičebním	2	v cvičebním	1
ve cvičení	27		0
ve cvičn*	7	v cvičných	1
ve Cvikově	2	v Cvikově	2

Table 132. Vocalisation of the preposition **v** before ⟨c⟩ in the contemporary part of the Czech corpus.

As there are only 8 more instances of collocations with the non-vocalised form of the preposition before a consonant cluster with an initial /c/, the total frequency of the vocalised form of the preposition before such clusters is 82%. It is however very difficult to trace any pattern in the distribution of the two forms of the preposition in this context.

For the context before [x] followed by a consonant, the data are somewhat easier to interpret.

ve chl*	12	4%	v chl*	313
ve chm*	6		v chm*	9
ve chr*	2	3%	v chr*	632
ve chř*	2		v chř*	11
ve chv*	2,194	99%	v chv*	21

Table 133. Vocalisation of the preposition **v** before ⟨ch⟩ followed by a consonant in the contemporary part of the Czech corpus.

Even though the number of instances is quite low in the majority of contexts, a pattern seems quite evident. The highest rate of vocalisation is found in the context of ⟨chv⟩ where the same phoneme is present both in the preposition and in the initial consonant cluster of the following word. It should especially be noted that before words based on the root {chvíl} *moment*, the vocalisation is practically obligatory. Only one single instance with the non-vocalised form of the preposition was found in the corpus, namely **v chvilkovou** in the *momentary*, whereas the vocalised form of the preposition was found in 2,156 instances in different collocational types with this root. The remaining collocations with ⟨chv⟩ exhibit the vocalised form of the preposition in 38 instances and the non-vocalised form in 20 instances, which still means a preference for the vocalised form at a rate of 66%.

In the context of ⟨chl⟩, the 12 instances of vocalisations were found before the root {chlév} *cowshed*, whereas 13 instances of the non-vocalised form occurred with the same root. Another 300 occurrences with the non-vocalised form occurred before other lexemes beginning with ⟨chl⟩.

With regard to ⟨chm⟩, five occurrences involved the vocalised form of the preposition before the root {chmel} *hops*, whereas only two instances involved the non-vocalised form before the same root. Another seven occurrences involved the non-vocalised form before other lexemes beginning with ⟨chm⟩.

With regard to ⟨chř⟩, two occurrences involved the vocalised form of the preposition before the root {chřtán} *throat*, whereas one instance involved the non-vocalised form before the same root. Another ten occurrences involved the non-vocalised form before other lexemes beginning with ⟨chř⟩.

The few remaining occurrences with the vocalised form of the preposition are all of an individual character and do not imply any distributional patterns.

The next context to investigate is before ⟨m⟩.

ve ma*	1	100%	v ma*	8,011
ve mc*	2		v mc*	2
ve md*	1		v md*	8
ve me*	1	0%	v me*	6,235
ve mě*	3,890	64%	v mě*	2,150
v mi*	1	0%	v mi*	12,542

v ml*	15	1%	v ml*	1,391
ve mn*	348	9%	v mn*	3,692
ve mo*	1	0%	v mo*	7,637
ve ms*	2		v ms*	1
ve mš	51		v mš*	4
ve mz*	49		v mz*	7

Table 134. Vocalisation of the preposition *v* before ⟨m⟩ in the contemporary part of the Czech corpus.

The highest number of occurrences of the vocalised form of the preposition is encountered before ⟨mě⟩, which is however pronounced [mɲɛ], hence these lexemes do phonetically begin with a consonant cluster. It turns out, however, that all the occurrences of the vocalised form in this context are found before two roots, namely in the collocation **ve mě** (10/0) *into me* and before the root {mňest} *town*. Furthermore, the occurrences involving the root {mňest} are limited to only five collocational types, namely collocations with the noun forms **ve městě** (3,485/180) *in the town*, **ve městech** (380/45) *in the towns*, **ve městečku** (13/157) *in the little town*, and with two different forms of the adjective **ve městsk*** (2/880) *in the town's*. The non-vocalised form of the preposition occurs 46 times in eight other collocational types involving the same root, e.g. **v Městci Králové** (10/0) *in Městec Králové* and **v měšťansk*** (22/0) *in the civic*, as well as more than 800 times in collocations with other lexemes beginning with ⟨mě⟩.

The second highest number of occurrences is found before ⟨mn⟩. All these occurrences involve the 1st person singular personal pronoun, i.e. **ve mne** (6/0) *into me* and **ve mně** (342/0) *in me*.

The only remaining substantial frequencies are found for collocations involving ⟨mš⟩, ⟨mz⟩ and ⟨ml⟩. Each of these initial consonant clusters exhibits frequent collocational types with one root only, namely **ve Mšeně** (46/1) *in Mšeno*, **ve Mšenu** (3/0) *in Mšeno*⁹⁶, **ve mzdách** (15/0) *in the salaries*, **ve mzdě** (10/0) *in the salary*, **ve mzdov*** (24/7) *in the salary's*, **ve mlejně**⁹⁷ (9/0) *in the mill*, **ve mlýně** (5/2) *in the mill*, **ve mlýnici** (1/10) *in the milling chamber*. Besides that, there are only two individual collocations with the vocalised form of the preposition before ⟨mš⟩, i.e. **ve Mšeckých** (1/2) *in Mšecké* and **ve mších** (1/0) *at the masses*. Before ⟨ml⟩, the non-vocalised form of the preposition occurs 1,391 times in a large variety of collocational types, among which the root {mlín} is present in the following collocational types: **v mlýnech** (2/0) *in the mills*, **v mlýnu** (2/0) *in the mill*, **v mlýnsk*** (7/0) *in the mill's*. Before ⟨mš⟩ the only

⁹⁶ One of many instances where Czech allows alternative endings in the same case.

⁹⁷ ⟨ej⟩ instead of ⟨ý⟩ reflects a colloquial pronunciation.

remaining instance of the non-vocalised form of the preposition was found in **v Mšeném** (1/0) in *Mšené*, whereas before ⟨mz⟩ no other instances occurred.

Among the remaining individual occurrences of the vocalised form of the preposition before ⟨m⟩, the following occurrences, involving complex consonant clusters, may be of interest: **ve mstě** (1/0) in *revenge*, **ve Mstěticích** (1/0) in *Mstětice* (countered only by **v mstivce** (1/0) *into revenge seekers*), **ve Mcelech** (2/0) in *Mcely* (but **v Mcelích** (2/0) in *Mcely*), **ve mdlobách** (1/0) in *unconsciousness* (but **v mdl*** (8/0) in *faint*).

Finally, it is interesting to note that only the non-vocalised form of the preposition occurs before ⟨mž⟩, namely **v mžiku** (41/0) in *the twinkle of an eye*.

The next context to investigate is before ⟨t⟩.

ve ta*	1	0%	v ta*	7,778
ve tb*	6		v tb*	33
ve te*/té*/tě*	4	0%	v te*/té*/tě*	29,886
ve tk*	9		v tk*	17
ve tl*	2	1%	v tl*	181
ve tm*	179	58%	v tm*	131
ve to*	1	0%	v to*	23,370
ve tr*	13	0%	v tr*	5,151
ve tř*	4,587	77%	v tř*	1,379
v tu*	1	0%	v tu*	6,969
v tv*	432	30%	v tv*	1,003

Table 135. Vocalisation of the preposition **v** before ⟨t⟩ in the contemporary part of the Czech corpus.

The two contexts with a majority of vocalised realisations of the preposition are before ⟨tř⟩ and ⟨tm⟩.

In the context of ⟨tř⟩, the two roots {tř} *three* and {tříd} *class, avenue* dominate the collocations with the vocalised form of the preposition. Before other lexemes, only 5 instances containing the vocalised form of the preposition were encountered, hence the non-vocalised form of the preposition is preferred almost without exception, e.g. **v Třebíči** (141/1) in *Třebíč*. Before the basic form of the numeral only the vocalised form occurs: **ve tři** (232/0) *at three, into three*. In other forms of numerals, based on the same root, the vocalised form dominates: **ve třech** (1131/3) in *three*, **ve třetí*** (1843/174) in *the third*, **ve třetině** (37/7) in *one third*, as well as in derivations beginning with **ve tři-/tří-** (723/348), whereas before further derivations the non-vocalised form dominates, e.g. **ve třetiligov*** (3/18) in *the third league's*. The vocalised form also clearly dominates in **ve třídě** (111/3) in *the class* and **ve třídách** (413/22) in *the classes*, but only the non-vocalised form occurs before derivations from this lexeme, e.g. **v třídn*** (6/0) in *the class's*.

In the context of ⟨tm⟩ all collocations involve the same root, namely ⟨tm⟩ *darkness*. In the singular **ve tmě** (147/0) *in the dark*, only the vocalised form of the preposition occurs, whereas in the plural **ve tmách** (3/4) *in the dark* both forms are almost equally frequent. Before the derived adjective **ve tmav*** (2/85) *in the dark*, the non-vocalised form dominates heavily. Before the geographical name in **ve Tmani** (27/39) *in Tmaň* the non-vocalised form dominates slightly, but before the corresponding adjective **v tmaňsk*** (3/0) *in Tmaň's* only the non-vocalised form occurs in the corpus. There were no other collocations involving the context of ⟨tm⟩.

In the context of ⟨tv⟩ the vocalised form occurs with a large number of different lexemes, but in most collocational types, the non-vocalised form dominates, e.g. **v tvrzení** (22/1) *in the claim*. There is also a large number of collocational types which only exhibit the non-vocalised form of the preposition. The highest frequency of the vocalised form occurs in the collocational types **ve tvaru** (234/4) *in the form*, **ve tvarech** (7/1) *in the forms*, **ve tváři** (121/22) *in the face*, **ve tvářích** (36/9) *in the faces*, all containing the locative case, whereas in **v tvář** (591/1) *into the face*, in the accusative, the non-vocalised form seems to be practically obligatory. A certain frequency of the vocalised form is also exhibited by **ve tvé/tvé**m (4/7) *in your*, **ve tvorbě** (7/82) *in the creation* and **ve tvrd*** (18/165) *in the hard*.

In the context of ⟨tk⟩, all the collocations contain the root ⟨tk⟩ *weave*. In all the collocational types there is a certain domination of the non-vocalised form of the preposition: **ve tkáni** (3/5) *in the tissue*, **ve tkáních** (4/7) *in the tissues*, **ve tkanině** (2/3) *in the fabric*, but only **v tkalcovně** (2/0) *in the weaving mill*. Similarly, in the context of ⟨tb⟩, all the collocations contain the name **Tbilisi**, and the non-vocalised form of the preposition seems to dominate: **ve Tbilisi** (4/32) *in Tbilisi*, **ve tbiliských** (2/1) *in Tbilisi's*.

Finally, in the context of ⟨tr⟩, one single collocation exhibits a quite exceptional frequency of vocalisation: **ve trojici** (8/13) *in the trio*, **ve trojících** (2/5) *in the trios*. A possible reason is influence from other lexemes containing forms based on the numeral **tři** *three*. Besides this single lexeme, there are only a couple of individual occurrences of the vocalised form of the preposition in this context, and the vast majority of collocations only occur with the non-vocalised form of the preposition.

In the remaining contexts the non-vocalised form of the preposition is absolutely dominant and the collocations occurring with the vocalised form of the preposition are all unique occurrences of a seemingly occasional character.

In the context of ⟨h⟩, there is a very high frequency of vocalisation before ⟨hř⟩ and ⟨hv⟩.

ve ha*	1	0%	v ha*	4,230
ve hb*	1		v hb*	8
ve he*	1	0%	v he*	1,486

ve hl*	13	0%	v hl*	7,868
ve hm*	7		v hm*	36
ve hn*	6	4%	v hn*	160
ve hr*	20	1%	v hr*	2,631
ve hř*	1,497	96%	v hř*	59
ve hv*	201	87%	v hv*	30

Table 136. Vocalisation of the preposition **v** before ⟨h⟩ in the contemporary part of the Czech corpus.

In the context of ⟨hř⟩, almost all occurrences with the vocalised form of the preposition represent a single collocational type: **ve hře** (1475/0) *in the game, in the play*. This is once again a noun with a stem that contains no vowel. There is also a number of collocational types with one or two occurrences each that do not have any corresponding occurrences with the non-vocalised preposition: **ve hřbetní** (1/0) *in the ridge's*, **ve hřbetu** (1/0) *in the ridge*, **ve hřbitov** (1/0) *into the churchyard*, **ve hřmotné** (1/0) *in the thunderous*, **ve hřmotu** (1/0) *in the thunder*, **ve hřšti** (2/0) *in the playground*. The derived adjective **ve hřbitovn*** (4/4) *in the churchyard's* exhibits an equal frequency of both forms of the preposition. All these lexemes, except for the last one, have in common that they begin with a cluster of three consonants. The last lexeme contains the same root as the noun in **ve hře**. The only other collocational types are **ve Hřensku** (10/27) *in Hřensko* and **ve Hředlích** (1/5) *in Hředle*.

For the context ⟨hv⟩, the situation is also limited to one single root, **hvězd-star**, and the two place names **ve Hvozdech** (2/0) *in Hvozdy* and **ve Hvoždanech** (1/1) *in Hvoždany*. The most frequent collocation contains the plural **ve hvězdách** (157/0) *in the stars*. Other forms with a predominance of the vocalised form of the preposition are **ve Hvěždě** (13/0) *in Hvězda*, **ve hvěždárně** (3/0) *in the observatory*, **ve Hvězdově** (1/0) *in Hvězdov*, **ve Hvězdně** (1/0) *in Hvězdno*, but with other forms the distribution is quite even **ve hvězdn*** (21/24) *in the star's*, **ve hvězdu** (2/3) *into the star*, and before the corresponding adverb only the non-vocalised form of the preposition was encountered: **v hvězdně** (1/0) *in the star-like*.

In the context of ⟨hr⟩, almost all collocations exhibit just a few occurrences, but interestingly enough, almost all of them are based on the root {hr} *play, game*: **ve hrách** (9/3) *in the games, in the plays*, **ve hru** (2/1) *into the game, play*, **ve hráče** (1/2) *into the players*, **ve hran*** (3/19) *in the played*. Of the remaining four collocational types, three single occurrences seem to be quite occasional, as they are much more frequent with the non-vocalised form of the preposition, whereas **ve hrozbách** (2/2) *in the threats* is equally frequent with both forms of the preposition.

In the context of ⟨hl⟩, ⟨hm⟩, ⟨hn⟩, ⟨hb⟩, ⟨ha⟩ and ⟨he⟩, almost all occurrences of the vocalised form of the preposition seem to be occasional, as they are

generally more frequent with the non-vocalised form. Only **ve hmotě** (4/2) *in the substance* and **ve hmotných** (2/1) *in the physical* are somewhat more frequent with the vocalised form, whereas **ve Hnátnici** (2/2) *in Hnátnice* is equally frequent with both forms of the preposition. The collocation **ve hbit*** (1/2) *in the alert* is also worth mentioning because of its unusual consonant cluster.

In the context of ⟨j⟩, very few collocational types were attested.

ve je*	3	0%	v je*	20,325
ve jm*	467	99%	v jm*	3

Table 137. Vocalisation of the preposition **v** before ⟨j⟩ in the contemporary part of the Czech corpus.

Apart from three occasional single occurrences before ⟨je⟩, the only context that systematically triggers vocalisation is before ⟨jm⟩. All those collocations contain the root {jmén} *name*: **ve jménu** (423/0) *in the name*, **ve jméně** (9/0) *in the name*, **ve jménech** (9/0) *in the names*, **ve jméno** (1/0) *into the name*, **ve jmenn*** (4/0) *in the name's*, **ve jmenování** (5/0) *in the naming*, **ve jmenovan*** (8/0) *in the named*, but **ve jemnovit*** (8/3) *in the nominal*.

The situation in the context of ⟨p⟩ is simpler than it might look at first glance.

ve pa*/pá*	1	0%	v pa*/pá*	19,873
ve pě*	1		v pě*	3
ve ph	147	76%	v ph*	46
ve pch*	1		v pch*	28
ve pl*	1	0%	v pl*	7,987
ve po*	1	0%	v po*	67,329
ve pr*	2,127	3%	v pr*	73,184
ve př*	14	0%	v př	35,429
ve ps*	1	0%	v ps*	339
ve pš*	2		v pš*	12
ve pt	3		v pt*	4

Table 138. Vocalisation of the preposition **v** before ⟨p⟩ in the contemporary part of the Czech corpus.

The collocational types heavily dominated by the vocalised form of the preposition are evidently restricted to two lexemes: **ve prospěch** (2,127/19) *in favour (of)* and **ve př** (9/0) *in the fight*. The occurrences of the vocalised form before ⟨ph⟩ are all instances of foreign names pronounced with an initial [f], e.g. **ve Phoenixu**, hence the vocalised preposition is expected. However, in the very same context, before ⟨ph⟩ pronounced as [f], the non-vocalised form also occurs 16 times, in addition to another 30 occurrences before ⟨ph⟩ pronounced

as [p], e.g. **v Phnompenhu** (25/0) in *Phnom Phen*. The reason for the use of the non-vocalised form of the preposition before (ph) pronounced as [f] might be that the writer was influenced by the spelling.

All the remaining occurrences before ⟨p⟩ are of an individual, possibly accidental character, but the collocational types **ve Pchjongjangu** (1/27) in *Pyongyang*, **ve Pfarzheimu** (1/0) in *Pfarzheim*, **ve pštrosech** (1/0) in *the ostrichs*, **ve Ptácké** (2/2) in *Pták's* and **ve Ptení** (1/0) in *Ptení* are worth mentioning because of the unusual consonant clusters involved.

In the context of ⟨l⟩, the four consonant clusters ⟨lh⟩, ⟨ln⟩, ⟨lv⟩ and ⟨lž⟩ trigger vocalisation of the preposition to a very high extent.

ve la*/v lá*	1	0%	v la*/lá*	3,002
ve lh*	110	96%	v lh*	4
ve ln*	13		v ln*	0
ve lo*	1	0%	v lo*	7,169
ve lv*	15		v lv*	1
ve lž*	8		v lž*	0

Table 139. Vocalisation of the preposition **v** before ⟨l⟩ in the contemporary part of the Czech corpus.

Among the more frequent collocational types are **ve lhůtě** (78/0) *within the time limit*, **ve Lhotce** (11/0) in *Lhotka*, **ve Lnářích** (11/0) in *Lnáře*, **ve Lvově** (9/0) in *Lvov*, **ve lži** (6/0) in *the lie*, **ve Lhoteckém** (6/0) in *Lhotka's*, **ve Lhase** (5/2) in *Lhasa*, **ve lhůtách** (4/0) *within the time limits*, **ve lví** (4/0) in *the lion's*, **ve Lhotě** (3/0) in *Lhota*, **ve lnářství** (2/0) in *the flax production*, whereas single occurrences were encountered of the following collocations: **ve lhání** (1/0) in *the lying*, **ve lhářku** (1/0) *into a liar*, **ve lhostejnosti** (1/0) in *the indifferentness*, **ve Lvovské** (1/0) in *Lvov's*, **ve lžici** (1/0) in *the spoon*, **ve lžích** (1/0) in *the lies* as well as the following occurrences with the non-vocalised form of the preposition: **v lhostejné** (1/0) in *the indifferent*, **v lhostejnost** (1/0) *into indifference*, **v lvici** (1/0) in *the lioness*.

In the context of ⟨r⟩, four consonant clusters trigger vocalisation in more than 90% of the instances.

ve rč*	2		v rč*	0
ve re*	1	0%	v re*	10,915
ve ro*	3	0%	v ro*	48,659
ve rt*	5		v rt*	0
ve rv*	29		v rv*	0
ve rw*	94		v rw*	3

Table 140. Vocalisation of the preposition **v** before ⟨r⟩ in the contemporary part of the Czech corpus.

The collocational types are: **ve rčení** (2/0) *in the saying*, **ve rtech** (2/0) *in the lips*, **ve Rtyni** (3/0) *in Rtyně*, **ve rvačce** (22/0) *in the fight*, **ve rváče** (1/0) *into a rowdy*, **ve rvačkách** (1/0) *in the fights*, **ve rvačku** (2/0) *into a fight*, **ve rvačky** (3/0) *into fights*, **ve Rwandě** (88/2) *in Rwanda* and **ve rwandsk*** (6/1) *in the Rwandan*.

In the remaining contexts, almost all collocations are represented by single occurrences of an occasional character before a single consonant followed by a vowel.

ve b*	3	0%	v b*	51,436
ve g*	5	0%	v g*	4,480
ve n*	13	0%	v n*	102,363
ve o*	3	0%	v o*	52,427
ve q*	1		v q*	73
ve ř*	1	0%	v ř*	8,326

Table 141. Vocalisation of the preposition **v** before ⟨b⟩, ⟨g⟩, ⟨n⟩, ⟨o⟩, ⟨q⟩ and ⟨ř⟩ in the contemporary part of the Czech corpus.

The only remaining occurrences before consonant clusters are **ve Nvotov*** (2/0) *in Nvot's*, **ve Bzenci** (1/10) *in Bzenec*. **ve Gstaadu** (4/9) *in Gstaad*, **ve Grenoblu** (1/38) *in Grenoble*.

There are also two interesting collocational types with the vocalised form of the preposition before a single consonant followed by a vowel, i.e. **ve níž** (3/4,821) and **ve němž** (1/6,890), both being forms of **jenž** *which*, a stylistically more literary relative pronoun than the standard **který** *which*; the vocalised form of the preposition possibly gives an even stronger stylistic effect.

This ends the investigation of the corpus data concerning the preposition **v(e)**.

8.1.2.3. z(e)

All instances of the vocalised form of the preposition **z(e)** were extracted from the corpus and sorted with respect to the first grapheme of the following word. Then all occurrences of the non-vocalised form occurring in the same contexts were counted for comparison.

During this process, quite a few instances were encountered that did not represent instances of the preposition, such as a misprinted diacritical sign, e.g. 'ze instead of **že** *that*, wrongly inserted spaces, e.g. **di vi ze** instead of **divize** *division*, **Ze** as a first name or as an abbreviation, **ze** incorrectly substituted for **za** *for* or **lze** *it is possible*. In five instances, collocations with **ze** were uninterpretable.

Further, the preposition **ze** occurred before the Roman numbers II, III, IV, XIV, XVI and XVII, in meta-discussions and before a misspelt word, e.g. **ze**

venského instead of the correct **ze slovenského** from the Slovak. All these occurrences of the form **ze** have been omitted from the following data.

For the non-vocalised form of the preposition, there were too many abbreviations and divided words to eliminate them all manually. They have therefore only been excluded from the data in the contexts where the number of non-vocalised occurrences was lower than the number of vocalised occurrences in the same context, e.g. before abbreviations such as **z STV** from *STV*, which would most probably be pronounced with an initial [ɛs]. There were also many instances where ⟨z⟩ did not represent the preposition, but constituted either an initial, a spelling mistake, a mistakenly detached part of another word, or part of an uninterpretable collocation. There were also collocations in which the word following the preposition was of German origin and the ⟨z⟩ would therefore be pronounced as [ts].

All the above mentioned instances have been disregarded in the data presented in the following table.

ze b*	36	0%	z b*	14,614
ze c*	137	1%	z c*	10,806
ze č*	2,204	18%	z č*	10,381
ze d*	3,258	11%	z d*	27,470
ze h*	896	6%	z h*	13,832
ze i*	1	0%	z i*	3,474
ze j*	75	1%	z j*	14,612
ze k*	1,477	7%	z k*	20,803
ze l*	128	1%	z l*	12,269
ze m*	711	3%	z m*	22,416
ze n*	7	0%	z n*	60,563
ze o*	1	0%	z o*	20,976
ze p*	58	0%	z p*	61,100
ze r*	59	0%	z r*	18,984
ze ř*	82	3%	z ř*	2,803
ze s*	47,155	100%	z s*	121
ze š*	4,151	80%	z š*	1,021
ze t*	2,222	6%	z t*	37,186
ze u*	1	0%	z u*	4,946
ze v*	5,214	17%	z v*	26,006
ze z*	19,197	100%	z z*	44
ze ž*	1,720	69%	z ž*	759
ze 1*	2	0%	z 1*	1,785
ze 2*	150	9%	z 2*	1,581
ze 3*	228	24%	z 3*	705
ze 4*	775	81%	z 4*	176

ze 5*	4	1%	z 5*	578
ze 6*	219	41%	z 6*	313
ze 7*	384	90%	z 7*	44
ze 9*	1	0%	z 9*	386

Table 142. Vocalisation of the preposition *z* in the Czech corpus.

The vocalisation before ⟨s⟩ and ⟨z⟩ is practically obligatory. A closer investigation of the collocations containing the non-vocalised form of the preposition reveals 10 instances with ⟨sh⟩ in English words and 9 instances of ⟨sch⟩ in German words. The remaining 102 instances occur to a high degree in the older part of the texts. In order to make the picture clearer, we will divide the data into older and contemporary texts, as well as separating ⟨š⟩ and ⟨ž⟩ followed by vowel and a consonant respectively.

ze s*	43,627	100%	z s*	25
ze z*	18,063	100%	z z*	6
ze šC*	2,503	100%	z šC*	2
ze šV*	1,348	58%	z šV*	965
ze žC*	158	99%	z žC*	1
ze žV*	1,189	65%	z žV*	645

Table 143. Vocalisation of the preposition *z* before sibilants in the contemporary part of the Czech corpus.

ze s*	9,528	99%	z s*	76
ze z*	1,134	97%	z z*	38
ze šC*	198	97%	z šC*	6
ze šV*	102	68%	z šV*	48
ze žC*	17	94%	z žC*	1
ze žV*	356	76%	z žV*	113

Table 144. Vocalisation of the preposition *z* before sibilants in the older part of the Czech corpus.

In contemporary Czech, vocalisation is practically obligatory before /s/ and /z/, as well as before /š/ and /ž/ followed by another consonant. However, if the sibilant is followed by a vowel, the vocalised form of the preposition is only slightly more common than the non-vocalised form. A comparison of the two tables also reveals that the rate of vocalisation has increased somewhat over time in the contexts that obligatorily trigger vocalisation, whereas it has fallen somewhat in the contexts where vocalisation is optional. In the contexts of ⟨š⟩ and ⟨ž⟩ followed by a vowel, no evident systematic differences in the use of the two forms of the preposition could be observed.

The vocalisations in the contemporary part of the corpus in the remaining

contexts are summarised in the following table.

ze č*	2,166	18%	z č*	9,629
ze v*	4,329	15%	z v*	23,787
ze d*	2,933	10%	z d*	25,147
ze h*	865	7%	z h*	12,320
ze k*	1,317	7%	z k*	18,552
ze t*	2,138	6%	z t*	32,379
ze m*	491	2%	z m*	20,660
ze ř*	45	2%	z ř*	2,663
ze l*	110	1%	z l*	11,037
ze c*	86	1%	z c*	9,898
ze j*	72	1%	z j*	12,990
ze b*	32	0%	z b*	13,573
ze r*	34	0%	z r*	17,764
ze p*	46	0%	z p*	57,682
ze n*	6	0%	z n*	54,855
ze i*	1	0%	z i*	3,344
ze u*	1	0%	z u*	4,658

Table 145. Vocalisation of the preposition *z* in the contemporary part of the Czech corpus.

The highest rate of vocalisation is triggered by words with an initial (č), and we will begin by investigating this context more closely.

ze čk*	9		z čk*	0
ze čl*	115	12%	z čl*	830
ze čt*	2,042	97%	z čt*	59

Table 146. Vocalisation of the preposition *z* before (č) in the contemporary part of the Czech corpus.

Only the three contexts (čk), (čl) and (čt) trigger vocalisation. The last is only represented by two roots, namely {čt} *read* and {čtr} *four*, whereas the first is represented by one single morpheme, {čkyň} *Čkyně*, and the second by three morphemes, {člen} *member*, {článk} *article*, {člun} *rowing-boat*.

Ze Čkyně (9/0) *from Čkyně* seems to be obligatory. The remaining collocational types are **ze členů** (103/570) *from the members*, **ze členství** (9/55) *from the membership*, **ze čtenářů** (8/11) *from the readers*, **ze čtenářek** (4/1) *from the female readers*, **ze čtenářsk*** (7/2) *from the readers'*, **ze čtení** (2/1) *from reading*, **ze čtyř** (1051/0) *from four*, **ze čtyř*** (275/28) *from four**, **ze čtvrtka** (111/0) *from Thursday*, **ze čtrvt*** (257/7) *from a fourth*/quarter**, **ze čtver*** (4/1) *from the square*, **ze čtverř*** (106/5) *from the quartet*, **ze čtrn*** (186/2) *from fourteen*. Two morphemes

occur only once: **ze článku** (1/34) *from the article*, **ze člunu** (1/2) *from the rowing boat*.

Within the contexts of the consonant clusters discussed, only one morpheme never triggers vocalisation with the non-vocalised form, i.e. **z člověka** (50/0) *from man*.

ze vč*	160	69%	z vč*	73
ze ve*	1	0%	z ve*	5,666
ze vh*	1		z vh*	7
ze vch*	2		z vch*	3
ze vk*	3	1%	z vk*	198
ze vl*	1	0%	z vl*	2,678
ze vm	2		z vm*	7
ze vn*	2	1%	z vn*	245
ze vp*	1		z vp*	1
ze vr*	4	0%	z vr*	1,360
ze vs*	516	100%	z vs*	1
ze vš*	3,112	98%	z vš*	79
ze vz*	515	98%	z vz*	8
ze vž*	0		z vž*	1

Table 147. Vocalisation of the preposition **z** before ⟨v⟩ in the contemporary part of the Czech corpus.

In the three contexts with the highest rate of vocalisation, i.e. before ⟨vs⟩, ⟨vš⟩ and ⟨vz⟩, there is an evident general tendency to vocalise. This tendency is strongest before ⟨vs⟩, where only one instance with the non-vocalised form of the preposition is present in the corpus. Before ⟨vz⟩, too, the instances with the non-vocalised form are few, but in one collocational type it occurs five times, namely **ze vzácn*** (19/5) *from the rare*. Before ⟨vš⟩ there are a few individual occurrences of the non-vocalised form as well as a few collocational types with several occurrences, i.e. **ze všedn*** (10/4) *from everyday's*, **ze všemocn*** (1/3) *from the omnipotent*, **ze všeobecn*** (56/58) *from the general*, **ze všudy** (2/4) *from everywhere*, but the general tendency in the vast majority of data is clearly to vocalise in this context. In the context of ⟨vž⟩, only one single occurrence is present in the corpus, namely **z vžitého** (1/0) *from the established*.

In the remaining contexts with a high rate of vocalisation, this is due to a high number of occurrences of the two collocational types **ze včera** (14/15) *from yesterday*, **ze včerejš*** (145/53) *from yesterday's*. The rest of the data relating to ⟨v⟩ consists of more or less individual occurrences.

ze dn*	771	66%	z dn*	404
ze dr*	134	6%	z dr*	2,224

ze dř*	157	36%	z dř*	276
ze dv*	1,859	58%	z dv*	1,334
ze dž*	8		z dž*	70

Table 148. Vocalisation of the preposition *z* before ⟨d⟩ in the contemporary part of the Czech corpus.

In the context with the highest relative frequency before ⟨d⟩, two lexemes produce all but two occurrences, namely the seemingly obligatory **ze dna** (136/0) *from the bottom*, **ze dne** (627/0) *from the day*, as well as the plural **ze dnů** (6/1) *from the days*, which is optional, but prevailing. Other lexemes based on the root {dn} *day* trigger vocalisation only occasionally.

In the context with the second highest rate, vocalisation is not obligatory in any collocational type. High frequencies are exhibited by **ze dvou** (1690/109) *from two*, **ze dveří** (47/1) *from the door*, **ze dvora** (31/2) *from the courtyard*, and **ze dvorů** (4/1) *from the courtyards*. Derivations from these lexemes, however, exhibit much lower frequencies, namely **ze dva*** (69/725) *from two**, **ze dvou*** (5/109) *from two**, **ze dvoj*** (11/328) *from double**, **ze dvork*** (2/2) *from the small yard*.

In the context of ⟨dr⟩, all occurrences except two are instances of the collocational type **ze druh*** (132/1507) *from the second*, where vocalisation is a less frequent, but possible option.

In the context of ⟨dř⟩, almost all occurrences are instances of the seemingly obligatory **ze dřeva** (152/0) *of wood*. The corresponding derived adjective, **z dřevě*** (24/0) *from the wooden*, does not exhibit any vocalisation, however. In the collocational type **ze dřezu** (1/1) *from the sink*, both forms were encountered only once.

Finally, the low frequency of vocalisation before ⟨dž⟩ may be readily explained if this sequence is considered to be one phoneme. Then the rate of vocalisation is 100% if this phoneme is followed by a consonant, but 0% if it is followed by a vowel. The only instances of vocalisation in the corpus are **ze džbánu** (7/0) *from the jug* and **ze Džmurových** (1/0) *from Džmur's*.

ze hr*	664	30%	z hr*	1,551
ze hř*	199	64%	z hř*	111

Table 149. Vocalisation of the preposition *z* before ⟨h⟩ in the contemporary part of the Czech corpus.

In the context before ⟨h⟩, three lexemes cover the majority of instances of vocalisation. The most frequent is **ze hry** (664/5) *from the game*, which once again is a noun lacking a vowel in its stem. The remaining types are **ze hřbitova** (18/1) *from the churchyard*, **ze hřiště** (166/49) *from the playground*, and **ze hřišť** (12/12) *from the playgrounds*. In the case of the last lexeme, vocalisation

is evidently much more frequent with the singular than with the plural. Three other collocational types with only individual occurrences were encountered, namely **ze hřbetu** (1/1) *from the ridge*, **ze hřebů** (1/0) *from the nails*, and **ze hřibů** (1/1) *from the boletuses*.

ze kb*	2		z kb*	1
ze kd*	2		z kd*	35
ze kř*	3	1%	z kř*	287
ze kt*	1310	81%	z kt*	300

Table 150. Vocalisation of the preposition **z** before ⟨k⟩ in the contemporary part of the Czech corpus.

In this context almost all instances of vocalisation are due to the interrogative and relative pronoun **ze kter*** (1309/266) *from which*, where the vocalised form of the preposition is evidently preferred. With the derivations **z kter*koli(v)** (28/1) *from whichever* and **z kter*ž(to)** (3/0) *from which* (exclusively relative), the non-vocalised form dominates, however. The only other lexeme beginning with ⟨kt⟩ does not trigger vocalisation, namely **z Ktory** (3/0) *from Ktora (Israel)*.

The data in the remaining contexts are very scanty, but a certain tendency to vocalise before ⟨kb⟩ is evident: **ze kbelíků** (1/0) *from the buckets*, **ze Kbelského** (1/0) *from Kbely's*, but **z Kbel** (1/0) *from Kbely*. Before ⟨kd⟩ and ⟨kř⟩, however, the tendency is not to vocalise, and the vocalisations encountered are only found in one collocational type each: **ze Kdyně** (2/0) *from Kdyně* and **ze křtu** (3/0) *from the baptism*, once again a root noun without a vowel, whereas the non-vocalised form of the preposition occurs in a number of different collocational types.

ze th*	1	1%	z th*	102
ze tl*	1	1%	z tl*	100
ze tm*	34		z tm*	20
ze to*	1	0%	z to*	12,887
ze tr*	2	0%	z tr*	4,026
ze tř*	2,097	71%	z tř*	855
ze tv*	2	0%	z tv*	407

Table 151. Vocalisation of the preposition **z** before ⟨t⟩ in the contemporary part of the Czech corpus.

In the above table, vocalisation before ⟨t⟩ only occurs to any significant degree in the contexts ⟨tm⟩ and ⟨tř⟩. In the first case, all occurrences but one are represented by the collocational type **ze tmy** (33/0) *from the dark*, whereas the derived adjective only marginally triggers vocalisation: **ze tmav*** (1/19) *from the dark*. In the second context, the root ⟨tř⟩ dominates: **ze tři** (1612/45) *from*

three, **ze třech** (14/0) *from three* (Common Czech), and **ze třetí*** (180/120) *from the third* exhibit a preference for the vocalised form of the preposition, whereas derivations involving **ze tři*/tří*** (254/396) *from three** and **ze třetí*** (14/45) *from third** exhibit a preference for the non-vocalised form.

Another important lexeme is **ze třídy** (16/2) *from the class; avenue* and the corresponding plural **ze tříd** (3/0) *from the classes; avenues*, whereas derivations exhibit a preference for the non-vocalised form: **z tříd*** (8/0) *from class**. The vocalised form of the preposition is also exclusively encountered in the collocations **ze třpytu** (1/0) *from the glitter*, **ze třpytiv*** (2/0) *from the glittering* and **ze třtiny** (1/0) *from the reed*. The non-vocalised form of the preposition is encountered before ⟨tř⟩ in numerous other collocations representing many different collocational types, e.g. **z Třince** (124/0) *from Trinec*.

ze mě*	191	12%	z mě*	1,345
ze mn*	108	9%	z mn*	1,139
ze ms*	15		z ms*	1
ze mš*	5		z mš*	2
ze mz*	159	98%	z mz*	4

Table 152. Vocalisation of the preposition *z* before ⟨m⟩ in the contemporary part of the Czech corpus.

Before ⟨m⟩, the majority of occurrences of the vocalised form of the preposition are found before the 1st person singular personal pronoun: **ze mě** (191/0) *from me*, **ze mne** (106/0) *from me*, **ze mně** (2/3) *from me*. In the last collocational type, the two instances with the vocalised form of the preposition are incorrectly spelled instances of the genitive case (correctly **mě** or **mne**), whereas the three instances with the non-vocalised form are correctly spelled instances of the dative, in which case the pronoun is not governed by the preposition, but by a following word which in turn is governed by the preceding preposition, e.g. **z mně dostupných informací** *from information accessible to me*. A second dominating and obligatory collocational type is **ze mzdy** (151/0) *from the wages*, whereas the corresponding derived adjective is used with both forms of the preposition: **ze mzdov*** (8/4) *from the wages'*. A third seemingly obligatory collocational type is **ze msty** (14/0) *from revenge*, but the adjective occurs once with the non-vocalised form: **z mstiv*** (1/0) *from the revengeful*. On the other hand, the agent noun occurs once with the vocalised form: **ze mstitelů** (1/0) *from the revengers*. Further, all instances of **ze mlejna** (12/0) *from the mill* are found in the film title **Princezna ze mlejna** *The Princess from the Mill*, which has a colloquial form. The standard **ze mlýna** (1/3) *from the mill* exhibits a preference for the non-vocalised form. Finally, **ze Mšena** (4/0) *from Mšeno* only occurs with the vocalised form, whereas the corresponding adjective **z mšenských** (1/0) *from Mšeno's* occurs just once, but with the non-vocalised

form. The name **ze Mšeckých** (1/1) *from Mšecké* occurs once in collocation with each of the forms of the preposition.

ze řa*/řá*	3	0%	z řa*/řá*	1,691
ze ře*	36	7%	z ře*	472
ze ři*/ří*	6	1%	z ři*/ří*	499

Table 153. Vocalisation of the preposition **z** before ⟨ř⟩ in the contemporary part of the Czech corpus.

In contrast with the great majority of other collocations with the vocalised form of the preposition, all instances of the vocalised form in the context of ⟨ř⟩ occur in a context where ⟨ř⟩ is directly followed by a vowel, hence the words do not begin with consonant clusters, as is normally the case after the vocalised form of the preposition. On the other hand, the rate of vocalisation is quite low and the only interesting collocational type, containing the vocalised form of the preposition in the majority of cases, is **ze řetězu** (35/14) *from the chain*, which only occurs in combination with the verbs **utrhnout se** *break loose* and **pustit** *let go* with the figurative meaning (*allow someone*) *to fly off the handle*.

ze lh*	21	z lh*	12
ze ln*	3	z ln*	0
ze lv*	16	z lv*	2
ze lž*	70	z lž*	3

Table 154. Vocalisation of the preposition **z** before ⟨l⟩ in the contemporary part of the Czech corpus.

In the context of ⟨l⟩, the most frequent, and seemingly obligatory, collocations are **ze lži** (60/0) *from the lie*, **ze lží** (7/0) *from the lies*, **ze Lvovic** (9/0) *from Lvovice*, **ze Lvova** (6/0) *from Lviv*, **ze lnu** (3/0) *from flax*, whereas the corresponding adjectives show variation: **ze lživ*** (2/2) *from the scurrilous*, **z lvovsk*** (2/0) *from Lviv's*. There are also two individual occurrences: **ze lvů** (1/0) *from the lions* and **ze Lžovic** (1/0) *from Lžovice*.

In the context of ⟨lh⟩, the variation is greater, but the vocalised form still dominates: **ze lhání** (3/0) *from lying*, **ze Lhotky** (10/1) *from Lhotka*, **ze Lhoty** (5/1) *from Lhota*, **ze lhůty** (2/1) *from the time limit*, whereas only the non-vocalised form occurs with the longer derivations **z lháštvi** (1/0) *from mendacity* and **z lhostejnosti** (6/0) *from indifference*.

In the context of ⟨c⟩, the majority of collocations turned out to contain words of foreign origin beginning with a ⟨c⟩ pronounced as [s]. All such occurrences have been disregarded in the following table.

ze ct*	6		z ct*	3
ze cv*	16		z cv*	8
ze cy*	1	0%	z cy*	234

Table 155. Vocalisation of the preposition **z** before ⟨c⟩ in the contemporary part of the Czech corpus.

All the instances of **ze** before ⟨ct⟩ are individual occurrences, but all contain words based on the root {ct} *honour*. The vocalised form is the only one found before nouns: **ze ctitele** (1/0) *from the admirer* (masc.), **ze ctitelek** (1/0) *from the admirers* (fem.), **ze ctnosti** (1/0) *from the virtue*, whereas the non-vocalised form occurs as well before adjectives: **ze ctíhodn*** (1/3) *from the honourable*, **ze ctížádostiv*** (1/0) *from the ambitious*, as well as **ze ctěnick*** (1/0) *from Ctěnice's*.

All the collocations before ⟨cv⟩ with more than a single occurrence involve the root {cvik} *exercise*: **ze cviku** (4/0) *from the practice*, **ze cvičení** (7/1) *from the exercise*, as well as the adjective **ze cvičn*** (2/3) *from the training's*, whereas the derivations **z cvičebných** (1/0) *from the training's* and **z cvičitelek** (1/0) *from the instructors* (fem.) only occur with the non-vocalised form. The vocalised form also occurs in **ze Cvikova** (1/0) *from Cvikov* and **ze Cvrčovic** (1/0) *from Cvrčovice*, as well as in **ze cvalových** (1/1) *from the gallop's*, but not in **z cvalu** (1/0) *from the gallop* and **z cválající** (1/0) *from the galloping*.

ze jm*	71		z jm*	19
--------	----	--	-------	----

Table 156. Vocalisation of the preposition **z** before ⟨j⟩ in the contemporary part of the Czech corpus.

In the context of ⟨jm⟩, the vocalised form of the preposition prevails, but with some variation. In the collocations **ze jmelí** (1/0) *from the mistletoe* and **ze jména** (32/0) *from the name*, only the non-vocalised form occurs, but with the plural **ze jmen** (11/1) *from the names* and the deverbal noun **ze jmenování** (4/1) the variation is only slight, whereas for the deverbal adjective **ze jmenovan*** (22/17), the distribution is more even.

ze bř*	32	9%	z bř*	344
--------	----	----	-------	-----

Table 157. Vocalisation of the preposition **z** before ⟨b⟩ in the contemporary part of the Czech corpus.

All the instances of vocalisation before ⟨b⟩ contain the lexeme {břeh} *shore*: **ze břehu** (20/6) *from the shore*, **ze břehů** (12/109) *from the shores*. The vocalisation is evidently optional, but still quite frequent. Just as with other lexemes, once again the unmarked singular form exhibits a higher rate of vocalisation than the more marked plural.

ze rč*	1	z rč*	0
ze rh*	1	z rh*	4
ze rt*	13	z rt*	0
ze rv*	7	z rv*	2
ze rw*	12	z rw*	28

Table 158. Vocalisation of the preposition *z* before ⟨r⟩ in the contemporary part of the Czech corpus.

The instances of vocalisation before ⟨r⟩ are quite few, but some patterns can still be observed. The very rare initial consonant cluster ⟨rč⟩ triggers vocalisation once in **ze rčení** (1/0) *from the idiom*. The non-vocalised form is not encountered before ⟨rt⟩ in the noun stem {rt} *lip* which does not contain any vowel, **ze rtu** (2/0) *from the lip*, **ze rtů** (10/0) *from the lips*, as well as **ze Rтынě** (1/0) *from Rтынě*. The vocalised form is also preferred before ⟨rv⟩: **ze rvaček** (1/0) *from the fights*, **ze rvačky** (4/0) *from the fight*, **ze rváčů** (1/2) *from the rowdies*, **ze Rvácova** (1/0) *from Rvácov*. There is also a considerable rate of vocalisation before ⟨rw⟩: **ze Rwandy** (12/24) *from Rwanda*, but once again this is not the case before the corresponding adjective **z rwandsk*** (2/0) *from the Rwandan*.

ze pě*	1	0%	z pě*	1,283
ze pl*	1	0%	z pl*	1,970
ze př*	2	0%	z př*	8,727
ze ps*	36	17%	z ps*	182
ze pš*	4		z pš*	13
ze pt*	2		z pt*	37

Table 159. Vocalisation of the preposition *z* before ⟨p⟩ in the contemporary part of the Czech corpus.

In the context of ⟨p⟩, a few collocational types deserve attention. Above all, it is clear that the vocalised form is preferred before the noun stem {ps} *dog* which does not contain any vowel: **ze psa** (6/1) *from the dog* and **ze psů** (24/1) *from the dogs*; this preference is less clear before the corresponding adjective: **ze psí*** (1/7) *from the dog's*. The derived geographical name in **ze Psár** (3/3) *from Psáry* also exhibits a medium rate of vocalisation, and a couple of forms of the verb **psát** *write* also exhibit a low rate of vocalisation, **ze psaní** (1/3) *from the writing*, **ze psaných** (1/0) *from the written*, whereas several other lexemes only occur with the non-vocalised form of this preposition. Another noun without any vowel in the stem triggers the vocalised form of the preposition in one occurrence: **ze pře** (1/0) *from the quarrel*, and three more geographical names trigger vocalisation once each: **ze Pšovky** (1/0) *from Pšovka*, **ze Ptení** (1/1) *from Ptení* and **ze Ptice** (1/0) *from Ptice*. Furthermore, two adjectives exhibit a low rate of vocalisation: **ze pšeničn*** (1/4) *from the wheat's*, **ze pštroší*** (2/6) *from the*

ostrich's. The remaining individual instances are not of any systematic interest, as the rate of vocalisation is so low that they must be interpreted as either mistakes or individual choices.

ze n*	6	0%	z n*	54,855
ze i*	1	0%	z i*	3,344
ze u*	1	0%	z u*	4,658

Table 160. Vocalisation of the preposition *z* before ⟨n⟩, ⟨i⟩, and ⟨u⟩ in the contemporary part of the Czech corpus.

All the occurrences before ⟨n⟩, ⟨i⟩ and ⟨u⟩ are of an individual character, and the rate of vocalisation is so low that they must be interpreted as either mistakes or individual choices that diverge from the predominant usage.

8.1.2.4. s(e)

Due to the fact that the vocalised form of this preposition is homonymic with the reflexive pronoun *se*, which is more than ten times more frequent than the vocalised form of the preposition, a quantitative investigation of the more than a million occurrences of the string ⟨se⟩ is impossible. As will be reported presently, the literature supports the view that the prepositions *s(e)* and *z(e)* behave in a very similar way. This similarity will be taken as a point of departure, and only a few more spurious collocational types will be investigated manually in order to support some of the conclusions drawn in the summary of the Czech data at the end of this chapter.

8.2. The Literature

With regard to Czech there seems to be even less written on the issue investigated than with regard to Slovak and Polish. The situation in the Czech Republic with regard to the publication of linguistic handbooks and dictionaries is similar to the one in Slovakia. Individuals and independent publishers seldom publish works of this kind. Instead they are generally published by the publishing house associated with the Academy of Sciences (Academia) or by the formerly state-owned specialist pedagogical publishing house (SPN). Codificational works are normally also approved by the Ministry of Education.

8.2.1. Dictionaries

For Czech, a modern corpus-based dictionary is still lacking. The newest dictionary editions are reprints of older works. The four volume *Slovník spisovného jazyka českého* (SSJČ'60), published in its first edition between 1960 and 1971, was reprinted in 1989, while the single volume *Slovník spisovné*

čestiny pro školu a veřejnost (SSČ'94) exists in four editions dated 1978, 1994, 2003, 2005, the two last editions containing only a few corrections in comparison to the second revised edition.

The major older dictionaries are Kott's seven volume *Česko-německý slovník* (Kott'1878) published between 1878 and 1893, the nine volume *Příruční slovník jazyka českého* (PSJČ'35), published between 1935 and 1957, and the single volume *Slovník jazyka českého* (SJČ'41), which, in its second, extended edition from 1941, contains no comments on the issue investigated other than for the preposition **k**.

8.2.1.1. **beze**

The older dictionaries do not comment on the usage of the vocalised form of this preposition, but SSJČ'60 says that it is used only before certain words beginning with a consonant cluster where the first consonant is ⟨s⟩, ⟨z⟩, or more seldom ⟨š⟩, e.g. **beze sporu**, **beze slova**, **beze zbytku**, **beze škody**, but sometimes also elsewhere, e.g. **beze všech dokladů** *without all documents*, **beze všeho** *without all; directly, without hesitation; of course*, **beze mne** *without me*. SSČ'94 states that this form is used before ⟨s⟩, ⟨z⟩, more seldom ⟨š⟩, and before consonant clusters. No examples are given.

Compared to the previous corpus findings, these statements are very vague. In the corpus examined, a number of other collocations also occurred with significant frequencies. Therefore, for comparison, the corresponding statistical data concerning the number of occurrences of the collocational types encountered with the vocalised form of the preposition, as well as the corresponding collocational types with the non-vocalised form of the preposition, have been retrieved from the Czech National Corpus⁹⁸.

beze cti	51		bez cti	26
beze jmen	128	65%	bez jmen	70
beze jména	667	67%	bez jména	331
beze mne	1,196	100%	bez mne	6
beze mě	1,603	100%	bez mě	2
beze slova	3,830	98%	bez slova	78
beze smyslu	609	73%	bez smyslu	227
beze sněhu	268	10%	bez sněhu	2,474
beze spánku	321	59%	bez spánku	223
beze spěchu	700	92%	bez spěchu	61

⁹⁸ Český národní korpus at <<https://www.korpus.cz/>>. Searches were made in 2014 in the whole SYN series of the synchronic written corpora with a size of 2,685 million words, using the KonText search engine available at <<https://kontext.korpus.cz/>>.

beze sporu	599	99%	bez sporu	7
beze stínu	133	15%	bez stínu	754
beze stop	666	88%	bez stop	77
beze stopy	3,695	94%	bez stopy	233
beze strachu	1,590	77%	bez strachu	475
beze studu	625	72%	bez studu	148
beze svědků	667	86%	bez svědků	109
beze světla	62	15%	bez světla	356
beze škody	550	82%	bez škody	120
beze švů	49		bez švů	11
beze všeho	829	90%	bez všeho	97
beze všech	355	40%	bez všech	550
beze vší	299	78%	bez vší	83
beze zbraně	809	92%	bez zbraně	67
beze zbraní	653	86%	bez zbraní	105
beze zbytku	36	37%	bez zbytku	61
beze změn	5,136	96%	bez změn	189
beze změny	5,558	88%	bez změny	743
beze ztrát	698	55%	bez ztrát	562
beze ztráty	341	4%	bez ztráty	9,063

Table 161. CNK data as of 25 May 2014 regarding collocations with **beze**.

As was already found based on the previously investigated corpus data, it is not possible to establish any systematic phonological principles for the vocalisation of this preposition. Instead, a certain number of collocational types exhibit high frequencies, both absolute and relative. These assumptions are confirmed by the CNK data. The individual collocational types therefore have to be treated as lexicalised entities. Besides the obligatory vocalisation in the collocational types **beze mne**, **beze mě** *without me*, it is especially common for adverbialised collocations expressing manner to exhibit high frequencies, of vocalisation, as in **beze sporu** *without dispute*, **beze slova** *without a word*, **beze změn** *without change*, **beze stopy** *without trace*, **beze spěchu** *without hurry*, **beze zbraně** *without weapon*, **beze všeho** *without hesitation* etc., whereas the preposition tends not to vocalise in the corresponding collocational types when used in their concrete, literal meaning.

8.2.1.2. **nade**

The older dictionaries do not comment on the usage of the vocalised form of this preposition, but SSJČ'60 says that the vocalised form is only used before certain words beginning with a consonant cluster consisting of two or more consonants. SSČ'94 states that this form is used before a consonant cluster. No examples are given.

Compared to the previous corpus findings, these statements are far too general. In the corpus examined, only very few collocational types occurred with any significant frequencies. Therefore, the corresponding statistical data concerning the previously encountered collocational types with the vocalised form of the preposition, as well as the corresponding collocational types with the non-vocalised form of the preposition, have been retrieved from the Czech National Corpus.

nade dnem ⁹⁹	112	49%	nad dnem	115
nade dveře	49	24%	nad dveře	159
nade dveřmi	432	56%	nad dveřmi	336
nade mne	64		nad mne	0
nade mě	61		nad mě	2
nade mnou	3,597	100%	nad mnou	6
nade Mží	0		nad Mží	0
nade vše	2,609	98%	nad vše	63
nade všechn.*	620	41%	nad všechn.*	907
nade vši	431	98%	nad vši	10
nade vší	456	82%	nad vší	101
nade vším	545	30%	nad vším	1,286

Table 162. CNK data as of 25 May 2014 regarding collocations with **nade**.

As was found based on our own corpus data, phonological principles do not trigger vocalisation of this preposition. However, a small number of collocational types exhibit high frequencies, which is also confirmed by the CNK data. These collocational types have to be treated as lexicalised entities. Apart from the obligatory vocalisation in the collocational types **nade mne**, **nade mě**, **nade mnou** *above me, over me*, there are a few adverbialised collocations which exhibit high frequencies, namely **nade vše**, **nade vším** *above all*, **nade vši pochybnost** *beyond all doubt*, as well as **nade vší pochybnost** with a common spelling mistake. In a few other collocational types, both forms of the preposition are almost equally frequent.

8.2.1.3. ode

Kott'1878 says that **ode** is used if the following word begins with a difficult consonant cluster, e.g. **ode dveří** *from the door*. The other two older dictionaries do not comment on the usage of the vocalised form of this preposition, but SSJČ'60 says that this vocalised form is only used before certain words beginning with a consonant cluster consisting of two or more consonants.

⁹⁹ Practically only collocations based on the neutre root {dn} *bottom*.

SSČ'94 states that this form is used before a consonant cluster. No examples are given.

Compared to the previous corpus findings, these statements are far too general. In the corpus examined, only very few collocational types occurred with any significant frequencies. The corresponding data concerning these collocational types have been retrieved from the Czech National Corpus.

ode dna	3,556	89%	od dna	421
ode dne	9,637	98%	od dne	172
ode dneška	23,138	96%	od dneška	1,041
ode dnů	24		od dnů	37
ode dveří	1,086	40%	od dveří	1,616
ode mě	10,166	99%	od mě	64
ode mne	9,199	99%	od mne	52
ode všeho	485	17%	od všeho	2,382
ode všech	729	7%	od všech	7,680
ode zdi	770	50%	od zdi	769

Table 163. CNK data as of 25 May 2014 regarding collocations with **ode**.

As was found based on our corpus data, phonological principles do not trigger vocalisation of this preposition. However, a small number of collocational types exhibit high frequencies, which is also confirmed by the CNK data. These collocational types have to be treated as lexicalised entities. Besides the obligatory vocalisation in the collocational types **ode mne**, **ode mě** *from me*, there are a few collocations which exhibit high frequencies, namely **ode dna** *from the bottom*, **ode dne** *from the day*, **ode dneška** *from today*, as well as **ode všeho** *of all* used without a following head word, and the longer phrase **ode zdi ke zdi** (637/101) *from wall to wall*. In the collocational type **ode dveří** *from the door*, both forms of the preposition are almost equally frequent, and the same is true in the longer phrase **ode dveří ke dveřím** (131/147) *from door to door*.

8.2.1.4. **pode**

Kott'1878 says that **pode** is sometimes used before consonants. The other older dictionaries do not comment on the usage of the vocalised form of this preposition, but SSJČ'60 says that the vocalised form is only used before certain words beginning with a consonant cluster consisting of two or more consonants. SSČ'94 states that this form is used before a consonant cluster. No examples are given.

Once again, these statements are far too general. In the corpus examined, only a couple of collocational types occurred with any significant frequencies. The corresponding data concerning these collocational types have been

retrieved from the Czech National Corpus.

pode dnem	101	36%	pod dnem	182
pode mě	68		pod mě	4
pode mne	52		pod mne	3
pode mnou	1,786	100%	pod mnou	1

Table 164. CNK data as of 25 May 2014 regarding collocations with **pode**.

As was found based on our own corpus data, phonological principles do not trigger vocalisation of this preposition. However, a couple of collocational types exhibit high frequencies, which is also confirmed by the CNK data. These collocational types have to be treated as lexicalised entities. Apart from the obligatory vocalisation in the collocational types **pode mě**, **pode mne**, **pode mnou** *under me*, only the collocation **pode dnem** *under the bottom* exhibits any significant vocalisation, but only at a lower frequency than the corresponding collocation with the non-vocalised form of the same preposition.

8.2.1.5. přede

The older dictionaries do not comment on the usage of the vocalised form of this preposition, but SSJČ'60 says that the vocalised form is only used before certain words beginning with a consonant cluster consisting of two or more consonants. SSČ'94 states that this form is used before a consonant cluster. No examples are given.

Once again, the statements are too general. In the corpus examined only a few collocational types occurred with significant frequencies. The corresponding data for these collocational types have been retrieved from CNK.

přede dnem	666	55%	před dnem	540
přede dvěma	2,556	2%	před dvěma	107,492
přede dveře	103	15%	před dveře	604
přede dveřmi	2,456	60%	před dveřmi	1,620
přede mě	710	100%	před mě	0
přede mne	550	100%	před mne	1
přede mnou	14,281	100%	před mnou	45
přede všemi	1,244	42%	před všemi	1,733
přede vším	198	31%	před vším	439

Table 165. CNK data as of 25 May 2014 regarding collocations with **přede**.

As was found based on our own corpus data, phonological principles do not trigger vocalisation of this preposition. However, a few collocational types

exhibit higher frequencies, which is also confirmed by the CNK data. These collocational types must be treated as lexicalised entities. Apart from the obligatory vocalisation in the collocational types **přede mě**, **přede mne**, **přede mnou** *in front of me, before me*, the vocalised and the non-vocalised form of the preposition is almost equally frequent in the collocational types **před(e) dnem** *before the day*, **před(e) dveřmi** *in front of the door* and **před(e) všemi** *in front of all*. In the meaning *right around the corner*, the collocational type **přede dveřmi** is clearly more frequent with the vocalised form of the preposition. The vocalised form of the preposition in the collocational type **přede vším** is above all due to the non-standard separated spelling of the lexicalised adverb **přede vším** *above all*.

8.2.1.6. přese

The older dictionaries do not comment on the usage of the vocalised form of this preposition, but SSJČ'60 says that the vocalised form is usually only applied in the collocation **přese všecko**, adding that in archaic and literary styles it also occurs before a cluster of two or more consonants. SSČ'94 states that this form is used in the collocation **přese všecko**.

Again, the statements are too general. In the corpus examined, only a small number of collocational types occurred with significant frequencies. The corresponding data for these collocational types have been retrieved from CNK.

přese mě	197	66%	přes mě	102
přese mne	140	87%	přes mne	21
přese vše	108	72%	přes vše	43
přese všecko	68	65%	přes všecko	36
přese všechna	16	1%	přes všechna	1,610
přese všechno	904	49%	přes všechno	942
přese všechnu	66	4%	přes všechnu	1,672
přese všechny	101	1%	přes všechny	6,955

Table 166. CNK data as of 25 May 2014 regarding collocations with **přese**.

As was found based on our own corpus data, phonological principles do not trigger vocalisation of this preposition. However, a small number of collocational types exhibit higher frequencies, which is also confirmed by the CNK data. These collocational types must be treated as lexicalised entities. The frequencies of vocalisation before the 1st person singular personal pronoun in the collocational types **přese mě**, **přese mne** *through me* are considerably lower than for other prepositions. It is also interesting that the more literary form **mne** triggers a higher rate of vocalisation than the stylistically more neutral **mě**. Further, the vocalised form of the preposition is only used with a high rate

of vocalisation in the adverbialised collocations **přese vše, přese všecko, přese všechno** *despite all*, whereas the non-vocalised form dominates in the same collocational types when used in other contexts.

8.2.1.7. zpode

The older dictionaries, and also SSČ'94, do not mention the vocalised form of this preposition, whereas in SSJČ'60 it is marked with a star, meaning that it occurs extremely rarely and is only included in the dictionary because it is used by some of the most prominent authors.

All the instances of the vocalised form of this preposition in CNK are presented in the following table together with the corresponding data for the non-vocalised forms in the corresponding collocational types.

zpode dveří	5	zpod dveří	34
zpode mě	2	zpod mě	0
zpode mne	3	zpod mne	0

Table 167. CNK data as of 25 May 2014 regarding collocations with **zpode**.

8.2.1.8. zpřede

Neither form of this preposition is mentioned in any of the five dictionaries and no instances were retrieved in CNK. Evidently, this preposition does not occur in Czech, in contrast with the Slovak **spred(o)** and the Polish **spzed(e)**.

8.2.1.9. poblíže

The older dictionaries give both forms as variants in the same entry without any comment on the usage of the individual forms. SSČ'94 does not mention the longer form.

Just as in the already investigated corpus data, the shorter form is absolutely dominant in CNK.

poblíže	547	1%	poblíž	91,758
---------	-----	----	--------	--------

Table 168. CNK data as of 25 May 2014 regarding **poblíž(e)**.

8.2.1.10. skrze

The five dictionaries investigated have sub-entries for the adverb and the preposition **skrz** respectively. The three older ones do not comment on the usage of the different forms. In SSJČ'60 the vocalised form is mentioned only in connection with the preposition and is said to be literary and archaic. In SSČ'94 the vocalised form is instead only mentioned as a variant of the adverb and its usage is said to be literary.

The observations already made in our own corpus are confirmed by data retrieved from CNK.

skrže	20,512	56%	skrz	15,867
-------	--------	-----	------	--------

Table 169. CNK data as of 25 May 2014 regarding **skrz(e)**.

8.2.1.11. díky

The older dictionaries give both forms as variants in the same entry without any comment on the usage of the individual forms. In SSČ'94 the entry **dík(y)** is listed after **diktovat** which suggests that the longer form **díky** is perceived as primary. The usage of the two forms is not commented upon.

The observations already made from our own corpus are confirmed by data retrieved from CNK.

díky	523,973	97%	dík	15,618
------	---------	-----	-----	--------

Table 170. CNK data as of 25 May 2014 regarding **dík(y)**.

8.2.1.12. kromě

Kott'1878 says that **kromě** is used before consonants, especially consonant clusters, whereas **krom** is used before a vowel in order to avoid hiatus. In SSJČ'60 the longer form **kromě** is listed as an entry in which **krom** is mentioned as a less frequent variant. In SSČ'94 the shorter form is not mentioned at all.

The observations already made in our own corpus are confirmed by data retrieved from CNK.

kromě	383,309	98%	krom	7,033
-------	---------	-----	------	-------

Table 171. CNK data as of 25 May 2014 regarding **krom(ě)**.

8.2.1.13. navzdory

In all the dictionaries investigated, both forms are given as variants without any comment on the usage of the individual forms.

The observations already made in the investigated corpus are confirmed by data retrieved from CNK.

navzdory	54,997	100%	navzdor	99
----------	--------	------	---------	----

Table 172. CNK data as of 25 May 2014 regarding **navzdor(y)**.

8.2.1.14. ke/ku

Kott'1878 says that **k** is replaced by **ke** or **ku** where euphony so requires, but also that this is not always maintained. PSJČ'35 says that the form **ku** is used before a few words beginning in **p**: **ku pomoci** *of help*, **ku příkladu** *for example*, **ku podivu** *surprisingly enough*, whereas SJČ'41 goes into more detail and says that **ke** is used according to current pronunciation habits: **ke cti** *to the honour*,

ke stu to a hundred, **ke dnu** to the bottom, **ke králi** to the king, **ke komu** to whom, **ke mně** to me, **ke lži** to the lie, **ke všem** to all...; sometimes both **k** and **ke** occur before the same word: **k/ke slovu** to the word, **k/ke stěně** to the wall, **k/ke svému** to one's own...; **ku** only occurs in fixed expressions before a labial consonant: **ku podivu**, **ku/k Praze** towards Prague, **ku příkladu**, **ku pomoci**, not **ku konci** towards the end, **ku kříži** to the cross, **ku králi**.

In SSJČ'60 the form **ke** is said to be used before ⟨k⟩, ⟨g⟩, usually before a cluster of three or more consonants and before a cluster of two consonants together with which **k** would be difficult to pronounce, whereas **ku** is used in some fixed collocations before an initial ⟨p⟩: **ku příkladu/kupříkladu**, **ku podivu/kupodivu**, **ku pomoci**, **ku prospěchu**, **ku Praze** and when expressing relationships, e.g. **3 : 1** read as **tři ku jedné**, **a : b** read as **á ku bé**.

SSČ'94 gives the same information in fewer words, stating that the form **ke** is used before ⟨k⟩, ⟨g⟩ and consonant clusters, whereas **ku** is used especially in the collocations **ku podivu/kupodivu**, **ku pomoci**, **ku Praze**, **ku prospěchu** for the benefit of, **ku příkladu/kupříkladu** and in mathematical expressions of the type **a : b** read out as **á ku bé**.

The general findings made based on the previous corpus investigation are well in line with the very general, broad claims made by the dictionaries. However, in addition to the already investigated corpus data, a smaller number of searches have been made in CNK for collocations which exhibited low absolute but high relative frequencies in our own corpus. Root nouns without a stable vowel in the root will be investigated separately in the summary of the Czech data.

ke břehu	2,560	89%	k břehu	323
ke břehům	446	38%	k břehům	719
ke cvič.*	814	81%	k cvič.*	188
ke členovi	8		k členovi	16
ke členu	13		k členu	6
ke členům	331	50%	k členům	335
ke dveřím	10,269	99%	k dveřím	72
ke džbán.*	17		k džbán.*	7
ke chvále	106	51%	k chvále	102
ke jmění	30		k jmění	8
ke jmenování	807	87%	k jmenování	118
ke jménu	786	91%	k jménu	76
ke jménům	31	10%	k jménům	287
ke lhaní	172	96%	k lhaní	8
ke lhůtám	32		k lhůtám	15
ke pštrosům	2		k pštrosům	6
ke třet.*	3,109	66%	k třet.*	1,598

ke vchodu	1,893	98%	k vchodu	44
-----------	-------	-----	----------	----

Table 173. CNK data as of 8 June 2014 regarding selected collocations with **ke**.

Based on a larger corpus, it becomes evident that a small number of collocations exhibit higher absolute frequency as well as a very high relative frequency.

8.2.1.15. se

The older dictionaries say nothing about the use of the vocalised form of this preposition. In SSJČ'60 the form **se** is said to be used before ⟨s⟩, ⟨z⟩, more seldom before ⟨š⟩, ⟨ž⟩, before clusters of three or more consonants, and also before clusters of two consonants together with which **s** would be difficult to pronounce. In SSČ'94 **se** is said to be used before ⟨s⟩, ⟨z⟩, more seldom before ⟨š⟩, ⟨ž⟩, and before consonant clusters before which **s** would be difficult to pronounce.

As already pointed out, this preposition cannot be quantitatively investigated due to its homonymy with the reflexive pronoun **se** *oneself*.

8.2.1.16. ve

The older dictionaries say nothing about the use of the vocalised form of this preposition. In SSJČ'60 the form **ve** is said to be used before words beginning with ⟨v⟩, and when needed also before clusters of two or more consonants together with which **v** would be difficult to pronounce. In SSČ'94 the form **ve** is said to be used before ⟨v⟩, ⟨f⟩, and before consonant clusters with which **v** would be difficult to pronounce.

The general findings based on the previous corpus investigation are well in line with the very general, broad claims made by the dictionaries. However, additional searches have been made in CNK for collocational types with low absolute but high relative frequencies in the previously investigated corpus, as shown in Table 174. Root nouns without a stable vowel in the root will be investigated separately in the summary of the Czech data.

ve dřevě	835	94%	v dřevě	54
ve dveřích	9,455	100%	v dveřích	7
ve džbán.*	205	91%	v džbán.*	21
ve Gstaadu	48	11%	v Gstaadu	403
ve hře	55,225	100%	v hře	60
ve chlév.*	592	58%	v chlév.*	422
ve Kbelích	886	89%	v Kbelích	111
ve kbelsk.*	5		v kbelsk.*	10
ve kbelíku	81	24%	v kbelíku	257
ve kter.*	296,885	93%	v kter.*	22,929

ve Kdyni	2,438	96%	v Kdyni	99
ve kdyň.*	60	38%	v kdyň.*	99
ve kdysi	12	2%	v kdysi	517
ve kdoví.*	1		v kdoví.*	41
ve křoví	807	23%	v křoví	2,628
ve Křtinách	922	94%	v Křtinách	61
ve kšeft.*	44	35%	v kšeft.*	81

Table 174. CNK data as of 12 June 2014 regarding selected collocations with **ve**.

Based on a larger corpus, it becomes evident that a small number of collocations exhibit higher absolute frequency as well as a very high relative frequency.

8.2.1.15. ze

The older dictionaries say nothing about the use of the vocalised form of this preposition. In SSJČ'60 the form **ze** is said to be used before ⟨z⟩, ⟨s⟩, more seldom before ⟨ž⟩, ⟨š⟩, before clusters of three or more consonants, and when needed also before a cluster of two consonants together with which **z** would be difficult to pronounce. In SSČ'94 the form **se** is said to be used before ⟨z⟩, ⟨s⟩, more seldom before ⟨ž⟩, ⟨š⟩, and before consonant clusters together with which **z** would be difficult to pronounce.

The general findings based on the previous corpus investigation are well in line with the very general, broad claims made by the dictionaries. However, additional searches have been made in CNK for collocational types with low absolute but high relative frequencies in our own corpus, as presented below. Root nouns without a stable vowel in the root will be investigated separately in the summary of the Czech data.

ze břehu	1,089	63%	z břehu	653
ze břehů	542	13%	z břehů	3,493
ze členů	4,306	25%	z členů	12,744
ze džbán.*	245	93%	z džbán.*	19
ze dveří	3,755	95%	z dveří	177
ze dvou	66,480	95%	z dvou	3,312
ze dvora	2,395	99%	z dvora	29
ze dvorů	97	69%	z dvorů	44
ze druh.*	7,336	11%	z druh.*	58,348
ze hřbetu	236	77%	z hřbetu	70
ze hřbitova	1,277	96%	z hřbitova	56
ze hřiště	10,956	82%	z hřiště	2,421

Table 175. CNK data as of 12 June 2014 regarding selected collocations with **ze**.

Based on a larger corpus, it becomes evident that a small number of collocations exhibit higher absolute frequency as well as a very high relative frequency.

8.2.2. Language Correctness

Language correctness handbooks are rather popular in the Czech Republic, but the issue investigated is seldom discussed in any detail. Some shorter mentions can however be found over the last hundred years.

Bartoš (1901: 2f.) says that prepositions ending in a consonant sometimes vocalise by adding an extra vowel **e** (**k** adds **u** before the labials **b** and **p**). It is impossible to enumerate the instances when this happens. If one has a good sense of the Czech language, one will always make the correct choice. In the 1870s voices were raised against the traditional prevailing habits, propagating the “euphonic” usage of forms such as **ve Praze** in Prague, **ve Brně** in Brno. This has led to many ugly vocalisations. The form **k** is losing ground since **ku** is increasingly used everywhere, even before words beginning with **u**, e.g. **ku účasti** to the participation, **ku obraně** to the defence. Similar exaggerated usage of **ve** can be also observed, e.g. **ve oku** in the eye, **ve úterý** on Tuesday.

Zenkl (1911: 136) says that **ke** is used before consonant clusters for better pronunciation and euphony, but never before a vowel or before **p**, whereas the form **ku** is used before **p** and sometimes before **b**, **m**, **v**. The use of **se**, **ve**, **ze** is not commented on here.

In a short note, the editors of *Naše řeč* (1917, I: 57) disapprove of collocations such as **ve městě** in the town, **ve bráně** in the gate.

Judging from these three publications, there seems to have been a growing tendency over a hundred years ago to apply the vocalised prepositions in new and unusual contexts, but most of the actual examples discussed have left almost no traces in modern Czech, with the exception of the very frequent collocation **ve městě**.

Šimek (1930: 204), in his review of a book translation, criticises improperly vocalised prepositions which he has encountered in collocations such as **ve městě** (3x) in the town, **ve své** (7x), **ve svém** (5x), **ve svých** (1x) in one's own, **ku své** (1x) towards one's own.

This author is also opposed to historically unmotivated vocalisations, but this time all his examples have become frequent in modern Czech usage, with the exception of **ku své**, where the vocalised form **ke** is used instead with very high frequency.

Kučař and Váhala (1959: 356ff.) state that before a vowel and before a single consonant differing from the consonant in the preposition, the asyllabic prepositions never vocalise. Before a single identical or similar consonant, they always vocalise. Before a cluster of two consonants where the first is different from the consonant in the preposition, asyllabic prepositions sometimes

vocalise, but the distribution in the present day language cannot be captured by any rule. It is the ease or difficulty with which the arising clusters may be pronounced that is decisive, but rhythmic considerations may also play a role. Most frequent are collocations such as **ve stodole** *in the barn*, **v stáří** *at old age*, **ve zpěvu** *in the singing*, **v spacím voze** *in the sleeping car*, **ve chvíli** *in the moment*, **v chvatu** *in a rush*, **v/ve spárech** *in the claws*, **ze dvora** *from the courtyard*, **z tvarohu** *of curd cheese*, **s ptákem** *with the bird*, **s ptaním** *with the asking*, **s tkanicí** *with the lace*, **se lžicí** *with the spoon*, **ze dne** *from the day*, **k studánce** *to the fountain*, **ke stolu** *to the table*. Therefore a great deal of variation can be observed in actual usage. The older generation prefers **v stáří** *at old age*, **s kterým** *with which*, but the younger more often **ve stáří**, **se kterým**. Generally, the younger generation uses vocalised forms more often than the older generation. If the second of the two consonants is **r**, **l** or **ř**, the non-vocalised form is normally used, but even in this context the younger generation sometimes applies the vocalised form, so that despite the regular **v slovech**, **k hradu**, forms as **ve slovech**, **ke hradu** occur very often. Before larger consonant clusters only the vocalised form is used.

The authors further point out that it is often not clear enough what is meant by “consonants similar to the consonant in the preposition”, and say that the corresponding voiced or voiceless consonant always leads to vocalisation. Before **š** and **ž**, however, the phonetically closely related prepositions **s** and **z** are usually not vocalised, even if the vocalised forms are not uncommon, e.g. **s širokým okrajem** *with a broad margin*, **s žábou** *with the frog*, **se šerpou** *with a sash*, **ze železa** *of iron*. Old habits are decisive here, but sometimes both forms are used, e.g. **s/se šaty** *with a dress*, **s/se ženou** *with the woman*. Often the final consonant of the preceding word also influences the choice of form of the preposition.

The authors also point out that the vocalised form **ku** was more common in older times and that it is only used in a few fixed collocations before the consonant **p**: **ku podivu/kupodivu** *astonishingly enough*, **ku příkladu/kupříkladu** *for example*, **ku Praze** *towards Prague*, **ku prospěchu** *for the benefit of* and **ku pomoci** *of help*.

It is also claimed by the authors that vocalisation of asyllabic prepositions before other somewhat similar consonants, such as **se/ze** before **ř** or **ke** before **h**, does not belong in good usage. Finally, vocalisation of syllabic prepositions ending in a consonant, such as **beze mne** *without me*, **beze slova** *without a word*, **nade vše** *above all*, **přede všemi** *in front of all*, occurs in very few collocations and there is practically no hesitation about its use.

From the previous corpus data, it becomes evident that some of the claims are not correct with respect to the present day situation. Vocalisation today occurs regularly and with a very high frequency before sibilant-initial consonant clusters such as **ve slovech**. It is also incorrect to say that the prepositions **s** and **z** would normally not vocalise before **š** and **ž**. Otherwise

the authors' claims are mainly confirmed by the corpus data.

Internetová jazyková příručka (IJP'2012) says that the reasons for vocalisation are partially historical, due to the evolution of Czech vowels, and partially phonetic, striving to facilitate pronunciation. Vocalisation still has not become stable and in certain contexts both forms occur. Therefore prevailing habits play a decisive role.

Asyllabic prepositions never vocalise before a word beginning with a vowel, but they always vocalise before a word beginning with an identical consonant like the one in the preposition itself. The only exception is **s sebou**. There is a strong tendency to vocalise before a similar consonant, i.e. **se** before **z-**, **š-**, **ž-**; **ze** before **s-**, **š-**, **ž-**; **ve** before **f-**; **ke** before **g-**. Here the phonetic factor is decisive. A few examples are e.g. **se zemí** *with the land*, **se ženou** *with the woman*, **se švestkou** *with the plum*, **s žaketem** *with the morning coat*, **ze stromu** *from the tree*. It is added that for **ve** vocalisation is absolutely dominant, and that for **ke** vocalisation is practically without exception. For **se** and **ze** it is noted that the non-vocalised form occurs to a small extent where it does not hinder pronunciation.

Asyllabic prepositions before other consonants exhibit variation. They do not vocalise before a single consonant. Before a cluster of two consonants they usually do not vocalise if the second of them is **r (ř)** or **l**, but they usually do vocalise before the clusters **tř**, **sl**, **dř**, **zr**, **zl**, e.g. **ve třech** *in three*, **ke třem** *towards three*, **ze třídy** *from the class*, **se třením** *with rubbing*, **ze dřeva** *of wood* (but **k dřevěné** *to the wooden*), **se slečnou** *with the miss*, **ve sluji** *in the grotto*, **ke slibu** *to the promise*, **ze slamníku** *from the straw mattress*, **se zrádcem** *with the traitor*, **ke zrodu** *to the birth*, **ve zrání** *in the maturation*, **ze zrady** *from betrayal*, **se zlodějem** *with the thief*, **ke zlatu** *to the gold*, **ve zlatě** *in the gold*, **ze zlosti** *of anger*. There are also other instances where vocalisation is preferred, e.g. **ve mlýně** *in the mill*, **ve hře** *in the game* (but **s hrou** *with the game*), but these are usually fixed collocations. Vocalisation is preferred if the second consonant is not **r (ř)** or **l**, and it is especially frequent when the second consonant is identical to the consonant in the preposition, e.g. **ve dveřích** *in the door*, **ve svém** *in one's own*, **ke skoku** *to the jump*.

Before a cluster of three consonants the asyllabic prepositions generally vocalise, e.g. **ke středu** *towards the middle*, **se vstupem** *with the entering*, **ve skladišti** *in the storehouse*, **ze vzpomínek** *of the memories*.

The preposition **k** also has the form **ku**, which is preserved in a few fixed collocations before a labial consonant, especially **p**, e.g. **ku příkladu** *for example*, **ku prospěchu** *for the benefit of*, and in expressions of relationship, e.g. **pět ku sedmi** *five to/versus seven*.

Finally, the syllabic prepositions are generally said not to vocalise, even if the following word begins with the same consonant as the final consonant of the preposition, e.g. **před domem**, **od dětí**. They only vocalise in certain specific contexts: before the pronouns **všechen** *all* and **já** *I* and in certain other

instances: **beze studu** *without shame*, **ode dveří** *from the door*.

The corpus data confirm the claims made, but the discussion in IJP'2012 about clusters of two consonants is unfortunately not very stringent, and the last examples of vocalisation of syllabic prepositions could have been chosen more carefully, among collocations exhibiting close to obligatory vocalisation.

8.2.3. Grammars

Reference grammars are more concerned with the issue under investigation than are dictionaries and language correctness handbooks. Most modern grammars have some notes on the usage of the vocalised forms of the prepositions.

Gebauer (1904: 16-18) is the first to give a sketch of the old Czech system of vowel–zero alternations, strictly following the principles for the fall of weak yers according to Havlík's rule. He then discusses changes in these rules. Firstly, modern Czech inserts a vowel between **s**, **z** and a sibilant, the only remnant of the old system being **s sebou** *with oneself*. No vowel is inserted before **c**, **č** or **ř**. In all the examples given, however, these three sounds are followed by a vowel. Secondly, old Czech collocations of the type **k králi** have been replaced by **ke králi** *to the king*. Thirdly, modern Czech prefers **ve** before **stř** and **skř**. In all these types the vowel is inserted because the old Czech collocations sounded hard or fierce (“*tvrdě*”) and the inserted vowel has an euphonic effect. Finally, Gebauer warns against exaggerated use of vowel insertion, e.g. before words beginning with a vowel, but also e.g. **ve koši** *in the basket*, **ve Klatovech** *in Klatovy*.

In the corpus data, no tendencies of the type represented by the last two examples were encountered.

Gebauer (1905: 12-14) says much the same, mentioning the forms **ve dne** *in the day* and **ke mně** *to me* as examples of historically correct forms. He then states that the correct usage of the vocalised prepositions is better preserved in the spoken language than in the written language, of which he gives examples such as **ve Praze**, **se přítelem**, **s lvem** instead of the correct **v Praze** *in Prague*, **s přítelem** *with a friend*, **se lvem** *with a lion* found in the spoken language. He recommends keeping to the spoken usage.

The examples of incorrect vocalisation mentioned in this grammar are practically absent from the corpus data.

Ertl (1920: 20) briefly repeats Gebauer's explanation of the usage of the vocalised form of prepositions according to Havlík's law. He gives the following examples of historically correct usage: **se psem** *with a dog*, **ve dne** *in the day*, **ze sna** *from the sleep*, **ode mne** *from me*, **beze snu** *without sleep*. He also repeats Gebauer's conclusion that correct usage of vocalised prepositions is better preserved in the general spoken language, which should be taken as guideline for usage; collocations such as **s lvem** *with a lion*, **s retů** *from the lips*,

v dnech *in the days*, **beze klamu** *without delusion* are to be avoided.

Ertl (1926: 42ff.) gives a very detailed explanation of Havlík's law with some traditional examples of vocalised prepositions. He then talks about instances when the historical law becomes disturbed through analogy and repeats that the spoken general language represents best usage. One should not write or pronounce e.g. **s lvem** *with a lion*, **pode skálou** *under the cliff*, **ve oku** *in the eye*, **ke svému** *to one's own*, **se kterým** *with which*, **ve příčině** *in the reason*.

In the corpus data, vocalisation of the asyllabic prepositions before a sibilant-initial consonant cluster, as well as before the cluster /kt/, exhibit a very high frequency.

Mašín (1928: 21, 111) makes only two very brief remarks. An intrusive **e** after prepositions is said to facilitate pronunciation, e.g. **ve škole** *at school*. It occurs after the prepositions **k, s, v, z, bez, nad, pod, před**, e.g. **ke škole** *towards school*, **se sestrou** *with the sister*, **ve dne** *in the day*, **ze země** *from the country*, **beze všeho** *without all; directly*, **nade vším** *over all*, **pode vším** *under all*, **přede vším** *in front of all; above all*.

Trávníček (1930: 19-24) mainly repeats Gebauer's (1904) explanation of Havlík's law and gives a few examples where the historical forms have been preserved, e.g. **ke dni** *to the day*, **ve dvou** *in two*, and some others where modern Czech has reinterpreted the stem of the word following the preposition, e.g. **k Chebu** *towards Cheb* for old Czech **ke Chbu** or **s dcerou** [st̩erou] *with the daughter* for old Czech **se dcerú**. He then discusses changes in the vocalisation rule, mainly repeating Gebauer (ibid.), but making one further clarification. He explicitly states that **ke** is used before words beginning with **k-**. Finally, he does not claim that vocalisation occurs because the clusters avoided sound hard or fierce, but rather because the clusters are difficult to pronounce, and the result of vocalisation is not euphonic, but simply easier to pronounce. Trávníček finally warns against exactly the same misuse of vocalisation as does Gebauer.

Gregor (1932: 34ff.) presents a quite different and very clear discussion of the issue. Having explained Havlík's law, he states that modern Czech departs from the historical principles through vocalisation of the prepositions **z** and **s** before **z, s, ž, š**, the preposition **v** before **v, f**, and the preposition **k** before words beginning with **k**. He then presents a number of historically correct vocalisations, e.g. **ve dne** *in the day*, **se rty** *with the lips*, **ve tmě** *in the dark*, **se mnou** *with me*, **ve dvou** *in two*, **ke všemu** *to all* and the archaic **se dcerou** *with the daughter*, which has been replaced by **s dcerou** in modern Czech. He finally discusses innovations in modern Czech, whereof the most interesting are **ke třem** *towards three* and **s třemi** *with three* instead of old Czech **k třem** and **se třmi**. He ends by remarking that the rules for vocalisation of prepositions has changed considerably over time, and many divergences have become acceptable in the literary language. However, it is not possible to use

vocalisation according to one's own preference, as it has evolved over centuries into the usage that can best be observed in the living language. Some common faults that should be avoided are **ve síle** in the strength, **od všech** from all, **v snách** in the dreams, **z štěstí** of luck.

Trávníček (1951: 46ff.) sees **blíž** closer and **skrz** through as instances of deletion of the final vowel of the same type as in e.g. **dřív** earlier or **hůř** worse, which are common in colloquial Czech instead of **dříve** and **hůře**, but also quite common in literary Czech.

He discusses the origin of parallel forms such as **ke/k svému** towards one's own, **ve/v svém** in one's own, and says that their origin is either based on common Slavic yers, which fell in Czech according to Havlík's law, or based on the insertion of a vowel into a consonant cluster in order to facilitate pronunciation. There follows a detailed explanation of Havlík's law and some examples of inconsistencies such as **se dvěma** or **s dvěma** with two, and he says that today it is difficult to recognize which alternations are based on actual yers and which ones are innovations. That **ke mně** to me and **ve mně** in me are historical forms, whereas **ke svému** towards one's own and **ve svém** in one's own are innovations, can only be distinguished by knowing historical grammar and etymology.

Trávníček states that the variation found in vocalisation in prepositional collocations is quite considerable, on the one hand between different collocational types such as **ke mně** to me, **k tobě** to you, **ve dne** in the day, **v noci** at night, but also within the same collocational type, such as **ke/k svému** towards one's own, **ve/v třídě** in the class. The form of the preposition depends on the following word. The vocalisation is then either based on inherited vocalisation of common Slavic yers, as in **ve vsi** in the village (< vъ vsi), or due to simplification of the pronunciation, as in **ve vodě** in the water (< vъ vodě) or **ve škole** at school.

In some instances, e.g. **k/ke svému**, **v/ve kterém** in which, the choice of form of the preposition is not stable and depends on rather individual factors. The degree of difficulty of pronouncing the consonant clusters arising from a preposition before a following cluster has become more powerful than the realisation of yers according to the historical sound law. This can be observed by studying different prepositions before the same word, e.g. **ke králi** to the king but **s králem** with the king, **se smíchem** with laughter but **k smíchu** to laughter. Therefore, vocalisation of prepositions can only be analysed based on the type of consonant cluster.

Prepositions never vocalise before a word beginning with a vowel. The asyllabic prepositions **k**, **s**, **v**, **z** vocalise more often than the syllabic prepositions. If the following word begins with the same voiced or voiceless consonant as the one in the asyllabic prepositions, then these prepositions are always vocalised, e.g. **ke králi** to the king, **se zájmem** with interest, **ve vodě** in the water, **ze sta** of a hundred.

The prepositions **s** and **z** sometimes vocalise before words beginning with **š** and **ž**, e.g. **s/se ženou** *with the wife*, **z/ze šatů** *of the clothes*. The vocalised form is more common than the non-vocalised, especially if the word preceding the preposition makes the cluster more difficult to pronounce, e.g. **přijdu zas se ženou** *I will come again with my wife*. Sometimes, however, the non-vocalised form is used as a rule, e.g. **s ševcem** *with the cobbler*.

Before a single consonant, prepositions as a rule do not vocalise. Only **ve městě** *in the town* occurs quite often alongside **v městě**. To this category also belong instances with a single consonant followed by a so-called syllabic **r** or **l**, e.g. **k srdci** *to the heart*, **v první** *in the first*.

If the word begins with a consonant cluster the prepositions sometimes vocalise, depending on the character of the consonant cluster. Apart from the already mentioned vocalisation of **k**, **s**, **v**, **z** before a similar voiced or voiceless consonant, the consonants **s** and **z** as a rule vocalise before **š** or **ž** followed by another consonant. Furthermore, the prepositions always vocalise before consonant clusters where **r**, **l**, **m** or **j** is the first consonant.

If the first in a cluster of two consonants is another consonant and the second is **r**, **l**, **m**, **n**, **ň** or **j**, the preposition generally does not vocalise. In some instances vocalisation is optional, e.g. **k/ku Praze** *towards Prague*, **k/ke sjezdu** *to the congress*, **k/ke hrobu** *to the grave*, **k/ke slávě** *to the glory*, **k/ke slovu** *to the word*, **v/ve slámě** *in the straw*, **v/ve sněhu** *in the snow*, **v/ve sluji** *in the grotto*, but only **ve Slezsku** *in Silesia*, **ve znaku** *in the sign*, **ve zlatě** *in gold*, and most often **ve prosěch** *for the benefit of*. Vocalisation in this context is less frequent before a complement than before a noun. Furthermore, monosyllabic words have a strong tendency to trigger vocalisation of a preceding asyllabic preposition, e.g. **ke dnu** *towards the bottom*, **ke dni** *to the day*, **ze dne** *from the day*, **ze dna** *from the bottom*, **ke snu** *to the dream*, **ve snu** *in a dream*, **ve snách** *in dreams*, **ve zlě**, **ve zlém** *in the evil*, **z/ze vrat** *from the gate*, **z/ze tmy** *from the dark*, **z/ze hry** *from the play*. The principle is that the non-vocalised form of the preposition is always possible whenever pronunciation is not difficult.

Before a cluster of two obstruents of which the first one is not identical or similar to the preposition itself, vocalised forms of the prepositions are more frequent than the non-vocalised forms, but there is a noticeable variation, e.g. **k světu** *to the world*, **k/ke světle** *towards the light*, **k/ke svému** *to one's own*, **ke škole** *towards the school*, **ke škodě** *to the detriment*, **ke skále** *towards the rock*, **ke skutku** *to the effect*, **ke dveřím** *towards the door*, **ve světě** *in the world*, **ve světnici** *in the chamber*, **ve svátek** *on the nameday*, **v/ve svém** *in one's own*, **ve chvíli** *in the moment*, **ve chvatu** *in haste*, **v květnu** *in May*, **v/ve tváři** *in the face*, **v tvé**, **v tvém** *in your*, **s/se dvěma** *with two*, **s/se kterým** *with which*, **se všech stran** *from all directions*, **se všemi** *with all*, **se ctí** *with honour*, **se čtyřmi** *with four*, **ze dveří** *from the door*, **ze dvora** *from the courtyard*, **ze čtyř** *from four*, **ze všech** *of all*, **ze vsi** *from the village*, **ze včerejška** *from yesterday*. Before the clusters **tř**, **dř**, **bř**, **př**, non-vocalised prepositions occur quite frequently, e.g. **k/ke třídě** *towards the class*,

k/ke dřevu *to the wood*, **k/ke břehu** *towards the shore*, **k/ke hřichu** *to the sin*, **s třídou** *with the class*, **s křídou** *with the chalk*, **s křížem** *with the cross*, **s přítelem** *with the friend*, **s břemenem** *with the burden*, **v/ve třídě** *in the class*, **v kříži** *in the cross*, **v přítomnosti** *in the presence*, **v břehu** *in the shore*, **v březnu** *in March*, **z/ze dřeva** *of wood*, **z dřívějška** *from earlier*, **z přátelství** *of friendship*, **z břehu** *from the shore*. Again, the non-vocalised form is more frequent before complements than before nouns, and the tendency for vocalised prepositions is strongest before monosyllabic words, e.g. **ke zdi** *towards the wall*, **ve zdi** *in the wall*, **ze zdi** *from the wall*, **ve dva** *into two*, **ve dvou** *in two*, **ve vsi** *in the village*, **ze vsi** *from the village*, **ve při** *in the fight*, **ve kři** *in the bush*, **ve tři** *into three*, **ve třech** *in three*, **ze čtyř** *of four*, the principle being that the non-vocalised form of the preposition is always possible whenever pronunciation is not difficult.

Before a cluster of three consonants where the first is not identical or similar to the preposition, the syllabic prepositions are most often vocalised. A non-vocalised preposition is quite rare, but it is certainly possible whenever it does not result in a cluster that is too difficult, e.g. **k zbláznění** *to becoming crazy*.

The syllabic prepositions never vocalise before a vowel, and they hardly ever vocalise before a single consonant, even if it is similar to the final consonant of the preposition. The only exception is **ode dávna** *since long ago*.

Vocalisation normally occurs with the pronoun **já** *I*, and sometimes before a consonant cluster where the first consonant is **r**, **l**, **m** or **n**, e.g. **beze lsti** *without any trick*, **bez/beze lhůty** *without time limit*, **bez/beze lžice** *without a spoon*, **bez mlýna** *without a mill*, **nad/nade mlýnem** *above the mill*, **pod/pode mlýnem** *under the mill*, **pode Lhotou** *beneath Lhota*, **přes mši** *before mass*, **přes rty** *over the lips*.

The pair **skrz/skrze** does not belong historically to this group of prepositions, as the short form has evolved from the long form by deletion of a final full vowel. The two forms are however applied according to the same principles as **bez/beze**.

There is generally no vocalisation before **kr**, **pr**, **sl**, **sm**, etc., where the second consonant is a sonorant. A few rare instances are **přede hrou** *before the game*, **bez/beze slova/slov** *without a word/words*, **ode dneška** *from today*, **ode dna** *from the bottom*, **přede dnem** *before the day*.

Before other consonant clusters, including those containing three consonants, the syllabic prepositions generally do not vocalise. A few collocations are frequent before forms of the pronoun **všechen** *all*; consider, however, **přes všecko úsilí** *in spite of all the effort*. Some instances of vocalisation occur if the first consonant in the cluster is similar to the last consonant of the preposition, e.g. **bez/beze strachu** *without fear*, **bez/beze studu** *without shame*, **beze změn** *without changes*, **od/ode dveří** *from the door*, **přede dveřmi** *in front of the door*, **nade dveřmi** *above the door*, **od/ode dvou** *from two* (but: **od dvanácti** *from twelve*), **pod/pode zdi** *under the wall*.

Syllabic prepositions vocalise far less frequently than the asyllabic prepositions. A syllabic preposition forms its own syllable, and the consonant cluster at the border between the preposition and the following word is thus divided into two syllables. Asyllabic prepositions have to be included in the same syllable as the initial consonant cluster of the following word. In old Czech, a non-vocalised asyllabic preposition would merge with the initial consonant of the following word, e.g. **k komu** *to whom* > **komu**. Such a merger however led to lack of clarity, because the preposition was actually lost, and therefore vocalisation was adopted.

Trávníček's discussion is probably the most thorough that exists. Most claims are confirmed by the corpus data. Only a few details are contradicted. It is not true that the asyllabic prepositions always vocalise before consonant clusters beginning with /r/, /l/, /m/ or /j/. Furthermore, saying that clusters with /r/, /l/, /m/, /n/, /ň/ or /j/ as the second element generally do not trigger vocalisation, is to simplify the picture too much. It is also quite incorrect to claim that the distribution of the forms **skrz/skrze** is the same as for **bez/beze**.

Šmilauer (1972: 33f.) states that asyllabic prepositions do not vocalise before vowels, nor before a single consonant differing from the consonant in the preposition. They do however vocalise before a single identical or similar (voiced, voiceless) consonant and sometimes in the type **ze života**. Before a cluster of two consonants, they normally do not vocalise if the second consonant is **r, ř** or **l**. Before other clusters of two consonants the usage varies, e.g. **k snatku** *to the wedding*, **v chvatu** *in hurry*, but **ve tváři** *in the face*, **ve Zbečně** *in Zbečno*. Before clusters of three consonants they normally vocalise, e.g. **ke hřbitovu** *towards the churchyard*, **se džbánkem** *with a mug*, **ve středu** *in the middle*, **ze mzdy** *of the wages*. The syllabic prepositions sometimes vocalise before **všechn** *all*, **mne**, **mnou** *me*, and sometimes before identical or similar consonants, e.g. **beze zvuku** *without a sound*, **beze slova** *without a word*. The preposition **k** also has the form **ku**, but only in the collocations **ku pomoci** *to help*, **ku Praze** *towards Prague*, **ku prospěchu** *for the benefit of*, **ku příkladu** *for example* and in mathematical expressions of the type **x : y** [iks ku ipsilon] *x (compared) to y*.

Havránek and Jedlička (1980: 143; 1981: 313) state that in order to facilitate pronunciation the asyllabic prepositions vocalise. This normally happens if the word following the preposition begins with an identical or similar consonant to the consonant in the preposition or with a cluster of three consonants. Before clusters of two consonants the prepositions sometimes vocalise and sometimes do not. The preposition **k** takes the form **ku** in the fixed collocations **ku podivu/kupodivu** *surprisingly enough*, **ku pomoci** *to help*, **ku prospěchu** *for the benefit of*, **ku Praze** *towards Prague*, **ku příkladu** *for example*, as well as in mathematical expressions of the type **a : b** read out as **á ku bé a** *(compared) to b*.

Also, syllabic prepositions ending in a consonant have vocalised forms,

but they are applied much less frequently than those of the asyllabic prepositions, e.g. **ode mne** *from me*, **pode mlýnem** *below the mill*, **beze slova** *without a word*, **beze všeho** *without all*, **nade vše** *above all*, **přeše všechno** *in spite of all*.

Mluvnice (1986) does not discuss vocalisation of prepositions.

Mluvnice (1995: 32, 345) states that some prepositions have two shapes. The vocalised form is suitable if the following word begins with the same or a similar consonant (a voiced or voiceless obstruent; an alveolar or palatoalveolar sibilant), with a cluster of two consonants if it does not contain /l/, /r/ or /ř/, or with a cluster of three consonants. The sensitivity of difficult clusters is individual, however, e.g. **k/ke tvé** *towards your*. Examples of vocalised syllabic prepositions mentioned are **nad/nade dveře** *above the door*, **ode dna** *from the bottom*, **beze všeho** *without all; directly*, **nade mnou** *over me*. It is added that the form of the preposition may also depend on other factors, such as rhythm. The preposition **k** also has the form **ku** before labials, but this form is literary and occurs above all in fixed collocations like **kupříkladu/ku příkladu** *for example*, **ku prospěchu** *for the benefit of*, **kupodivu** *surprisingly enough*, and also with digits to express mathematical relations like **pět ku šesti** *five to six*.

School and university grammar textbooks touch very lightly on this subject and their claims are not very stringent.

Pauliny (1972) in his textbook for Slovak students says that the asyllabic prepositions (**s**, **z**, **v**, **k**) and the syllabic prepositions ending in a consonant (**od**, **nad**, **pod**, **bez**, **přes**, etc.) vocalise under certain conditions. If the word following the preposition begins with the same or a similar consonant to the one at the end of the preposition, or with a cluster of three consonants, the preposition occurs in its vocalised form, e.g. **ze závisti** *with envy*, **ve vodě** *in the water*, **se škodou** *with the damage*, **ve směru** *in the direction*, **ke komu** *to whom*, **beze všeho** *without all; directly*, **ode dvou** *from two*, **přede mnou** *before me*, **pode mlýnem** *beneath the mill*. The preposition **k** also has the form **ku** in certain fixed collocations: before labial consonants in **ku podivu/kupodivu** *surprisingly enough*, **ku pomoci** *to help*, **ku prospěchu** *for the benefit of*, **ku Praze** *towards Prague*, **ku příkladu/kupříkladu** *for example*, and in mathematical expressions like **á ku bé** (a : b).

Melichar and Styblík (1981: 147) say that for better pronunciation the asyllabic prepositions **s**, **z**, **v**, **k** are sometimes vocalised. In a few fixed collocations before words beginning with **p** the preposition **k** has the vocalised form **ku**: **ku prospěchu** *for the benefit of*, **ku pomoci** *to help*, **ku Praze** *towards Prague*, **ku příkladu** *for example*.

Styblík et al. (1992: 88) say that the asyllabic prepositions **s**, **z**, **v**, **k** sometimes have the vocalised shape **se**, **ze**, **ve**, **ke** in order to facilitate pronunciation. This is usually so before a word beginning with an identical or a similar consonant to the consonant in the preposition, as well as before a

word beginning with a cluster of three consonants. Before a cluster of two consonants sometimes the vocalised and sometimes the non-vocalised form is used, e.g. **ve stole** in the table, **v starostech** in the worries. The preposition **k** also has the form **ku**, but only in the collocations **ku příkladu/kupříkladu** for example, **ku podivu/kupodivu** surprisingly enough, **ku pomoci** to help, **ku prospěchu** for the benefit of, **ku Praze** towards Prague and in mathematical expression such as **a : b (á ku bé)**. Syllabic prepositions ending in a consonant (**bez, od, pod, nad, před, přes**) also sometimes vocalise in order to facilitate pronunciation: **beze mne** without me, **ode mne** from me, **nade mnou** over me, **pode mnou** under me, **přede mnou** before me, **přeše všechno** in spite of all.

Novotný et al. (1992: 63) say that some prepositions vocalise in order to facilitate pronunciation before a word beginning with an identical or a similar consonant. The preposition **k** also has the form **ku**, but only in fixed collocations before the labial consonant **p**: **ku podivu/kupodivu** surprisingly enough, **ku pomoci** to help, **ku Praze** towards Prague.

Hlavsa et al. (1996: 95) only mention that the asyllabic prepositions **s, z, v, k** sometimes have the syllabic shape **se, ze, ve, ke** in order to improve pronunciation. The preposition **k** also has the form **ku** in the collocations **ku podivu** surprisingly enough, **ku pomoci** to help, **ku prospěchu** for the benefit of, **ku Praze** towards Prague, **ku příkladu** for example.

When it comes to grammars intended for foreigners studying Czech, amazingly little is written about the issue investigated, and the directions given are so vague that they are not of much help.

Mazon (1921: 30) says that an **-e** is added to several prepositions in order to split consonant clusters that would be difficult to pronounce, e.g. **ke stolu** towards the table, **ve středu** on Wednesday. The preposition **k** adds **u** instead of **e** before words beginning with **p**, e.g. **ku podivu** surprisingly enough, **ku předu** forward, **ku příkladu** for example.

Bauernöppel et al. (1970: 17) only mentions **ode mne** from me as a general example of vowel insertion.

Vintr (1994: 100) says that the asyllabic prepositions **k, s, v, z** (more seldom also **nad, bez, od, pod, přes, skrz**) are vocalised when the following word begins with the same or a similar consonant to the one in the preposition, or with a consonant cluster. Apart from very obvious examples, only **beze změny** without change and **nade dveřmi** above the door are given. It is also added that **k** is vocalised as **ku** before words beginning with **p-**, e.g. **ku Praze** towards Prague, **ku pomoci** to help, **ku příkladu** for example.

Janda and Townsend (2002: 11) say that vowel epenthesis is regular only before the oblique forms of the 1st person singular personal pronoun, and somewhat less regular before forms of **všechen** all. The presence of the same consonant or a voiced/voiceless counterpart makes epenthesis mandatory. Epenthesis is optional when the prepositions **s** and **z** are followed by the similar consonants **š** and **ž**, e.g. **s/se školou** with school, **s/se ženou** with the

woman. The preposition **k** can have **u** as its epenthetic vowel in certain set expressions with nouns beginning with **p-**, e.g. **ku podivu** *surprisingly enough*. Otherwise the normal **e** appears, e.g. **k/ke stolu** *towards the table*, **ke mně** *to me*.

Naughton (2005: 203) says that some asyllabic prepositions often have an appended **e**. Generally this occurs before a consonant cluster, before an identical consonant or before the voiced or voiceless counterpart of the consonant in the preposition. Some prepositions ending in consonants add **e** before **mě, mne, mně** and forms of **všechno**.

Obuchova (2010: 195f.) says that the monosyllabic prepositions **k, v, z, s** receive an extra **e** before words beginning with a consonant cluster, as well as before the consonants **k, v, z, s** and other consonants produced in a similar manner. Some of the more interesting examples are **ve tvaru** *in the shape*, **ve městě** *in the town*, **v společnosti** *in the society*, **ve státech** *in the states*, **ve sněhu** *in the snow*, **ve škole** *in the school*, **ke vzorům** *to the patterns*, **ze třídy** *from the class*. The prepositions **od, před, nad, bez** also exhibit alternations, e.g. **ode mne** *from me*, **přede mnou** *before me*, **nade mnou** *above me*, **beze mne** *without me*.

In CNK the vocalised form of the preposition is however circa 1000 times more frequent than the form given in the example, **v společnosti**.

8.2.4. Linguistics

There are actually only a few linguistic articles on various aspects of the matter investigated, and some shorter passages in a couple of monographs. As a rule, the corresponding alternations in prefixes get more attention.

Ryba (1934) states that linguistic recommendations and people's feeling for the vocalisation of prepositions are often in conflict, and that linguists themselves often use vocalised forms which they severely criticise, especially **ve** and **ke** before consonant clusters beginning with a sibilant. He gives a short historical background saying that euphony had been taken as the main line of guidance until A. Havlík pointed out the historical regularities that could be found in old Czech, where instances like **v svém** *in one's own*, **s svatým** *with the holy*, **z slov** *from the words*, **k králi** *to the king* or **v vás** *into you* are the norm. Ryba questions whether this old Czech usage should be allowed to influence modern standard Czech. Havlík, being the discoverer of the regularities in old Czech, reportedly defends the application of these principles even in modern Czech, stating that people speaking other languages, e.g. Italian, have no problem pronouncing long consonants, so why should it be a problem in Czech?

Ryba however argues that the unvocalised prepositions in such collocations belong irretrievably to the distant past. Since then analogical changes have altered the situation, and a new modern Czech usage has come to prevail. Actually some such discrepancies were already to be found in old

Czech, and even Havlík himself has mentioned a few. Ryba states that the new usage has a sound foundation in spoken common Czech as well as in the dialects, and that only very little variation can be observed between individual dialects. The new types of vocalisation can be found both in good prose and in the writings of outstanding bohemists. It is therefore time for the linguistic recommendations to adjust. He finally adds that there is an extra value added to the language through instances where a choice exists between the vocalised and the non-vocalised form of a preposition. This systemic instability can be used, for instance, for rhythmical reasons in poetry, to achieve stylistic differences between authors, or even to make a deliberately artificial impression, as Čelakovský does with the phrase **ve bílá ěadra** into *the white breasts*.

Haller (1934) discusses the differences between the available sources and suggests that Havlík's idea of following the historical principles is not realistic, preferring Gebauer's recommendation of following the usage in the common spoken language, where analogical changes have been introduced, for greater clarity and a more comfortable pronunciation. It is problematic, however, that common spoken language usage has not been studied and described. As an example, he gives Ertl's (1926:44) claim that **ke svému** *to one's own*, **se kterým** *with which*, **ke svatbě** *to the wedding*, **ke strýčkovi** *to the uncle*, **ze hmoty** *of the material*, **ve svém** *in one's own* are inappropriate. Instead, Haller claims that most speakers use such collocations regularly, and he says that the principles ought to be investigated, formulated, and adapted as a norm. Therefore, he has excerpted all the relevant prepositions on one hundred pages each in ten novels by ten different distinguished modern authors and found that the agreement between the individual authors was almost complete.

The asyllabic prepositions never vocalised before a vowel. Before a consonant identical to the one in the preposition itself, they always vocalised, with the single exception of **s sebou** *with oneself*.

Before a word which begins with a similar consonant to the one in the preposition, i.e. **s z-**, **s ž-**, **s š-**, **z s-**, **z š-**, **v f-**, **k g-**, the prepositions almost always vocalised. It should be noted for some reason that the context **z ž-** is not mentioned. The prepositions **s**, **z** sometimes did not vocalise before **š-**, **ž-**, but this usually only happened when the initial sibilant was followed by a vowel, e.g. **s ženami** *with the women*. In this type of context the non-vocalised form occurred in about 30% of instances.

Before a word beginning with any other consonant, vocalisation sometimes occurred, sometimes not. Here, the asyllabic prepositions never vocalised before a single consonant or before **mě-**. The collocation **ve městě** did not occur.

Before a cluster of two consonants the asyllabic prepositions did not normally vocalise if the second one was **r (ř)** or **l**. Only in about 10% of the instances was a vocalised preposition used, especially before **tř**, **dř**, **sl**, **zr**, **zl**, as

well as in some other instances, e.g. **ve třech** *in three*, **ve třetím** *in the third*, **ze třetí** *from the third*, **se třapcem** *with the tassel*, **ve třídu** *into the class*, **ke třídě** *towards the class* (but **s třemi** *with three*, **s třetím** *with the third*, **v třetím** *in the third*, **v třikrát** *in three times*), **ve Dřevíči** *in Dřevíč*, **ze dřeva** *of wood*, **ze dříví** *of timber* (but **v dřevíčském** *in the ... of Dřevíč*, **z dřeva** *of wood*, **k dřevěné** *to the wooden*), **ve slečnách** *in the misses*, **ve sluji** *in the grotto*, **ve slunci** *in the sun*, **ve slovenských** *in the Slovak*, **ve Slezsku** *in Silesia*, **ve službě** *in the service*, **ve slamníku** *in the straw mattress*, **ke slamníku** *to the straw mattress* (but **v slovo** *into the word*, **k slečnám** *to the misses*, **k sluchu** *to the hearing*), **ve zraky** *in the eyes*, **ve zlé**, **ve zlém** *in the evil*, **ke Zlíčku** *towards Zlíčko*, **ze břehů** *from the shores*, **ve mlýně** *in the mill*, **ze mlejna** *from the mill*, **ke hrobu** *towards the grave*, **ve hře** *in the game*, **ke hře** *to the game*, **ze vrat** *from the gate*, **ze vrátek** *from the wicket*. He interprets the reasons for vocalisation sometimes as reflexes of the historical regularities, sometimes as analogical innovations in order to avoid uncomfortable consonant clusters, and sometimes both at the same time.

If any other consonant is the second one in a cluster not beginning with an identical or similar consonant to the one in the preposition, the preposition is sometimes vocalised, sometimes not, but a striving to avoid complicated consonant clusters and to facilitate pronunciation can be observed. The asyllabic prepositions are vocalised if the second of the two consonants is the same as the preposition itself, e.g. **ve dvoře** *in the courtyard*, **ke škole** *towards school*. The non-vocalised form occurs in only about 20% of the instances. It can also be observed that the prepositions are generally vocalised whenever the result would otherwise have been a cluster which does not occur tautosyllabically in the language, e.g. **ke rtům** *to the lips*, **ve jménu** *in the name*, **ve lhůtách** *in time-limits*, **se čtyřmi** *with four*, **se ctí** *with honour*, **ze včerejška** *from yesterday*, **ze kterých** *from which*. Many such instances also constitute fixed collocations, e.g. **se mnou** *with me*, **se všech stran** *from all directions*, **se rty** *with lips*, **ze tmy** *from the dark*, **ze čteného** *from the read*, **ze vsi** *from the village*, **ze psů** *of the dogs*, **ve dne** *in the day*, **ve jménu** *in the name*, **ke psí boudě** *towards the kennel*, **ke cti** *to the honour*, **ke lvovi** *to the lion*. The greatest freedom of choice is found with the preposition **v** before clusters beginning with a sibilant consonant, where the frequency of vocalisation is close to 50%. This is explained as an effect of two possible ways to pronounce the preposition, either carefully with a complete labio-dental closure, which is more difficult and leads to vocalisation, or more hastily, only approaching the teeth to the lower lip. The same principle could also explain parallel instances such as **v/ve mnohém** *in many*, **v/ve kterých** *in which*, **v/ve hmotě** *in the material*. When difficulties do not arise the preposition is left unvocalised, which is the case of **k** before a sibilant or **tv-**, e.g. **k světu** *to the world*, **k tvému** *to your*. The prepositions **s**, **z** also have an ability to connect with many consonant clusters, e.g. **s knedlíkem** *with the knödel*, **z kterých** *of which*, **s tmou** *with the dark*, **s ptákem** *with the bird*, **z vnitra** *from the inside*, **s chtivostí** *with greed*, **z hnízda**

from the nest, even though some of the authors sometimes vocalise in such contexts, e.g. **ke zboru** to the choir, **ke svému** to one's own, **ke schodům** towards the stairs, **ke stolu** to the table, **ke tváři** towards the face, **ke stěně** towards the wall, **se kterou** with which, **ze kterých** of which, **ze dvora** from the courtyard, **ze dveří** from the door, **ze dvou** of two, **ze dvanácti** of twelve. Consequently, in many instances vocalisation must be considered optional. Often the context of the preceding word plays a decisive role, so that if the preceding word ends with a consonant, the tendency is to vocalise.

Before a cluster of three or more consonants the asyllabic prepositions vocalise. Only if the last of the consonants in the cluster is **r** (**ř**) or **l** and if the same consonant as the one in the preposition itself is not present in the cluster, is the non-vocalised form found, in about 10% of the instances, e.g. **v strýcově...** in the uncle's..., **v strašně...** in the terribly..., **v Sklenařicích** in Sklenařice, **k zbláznění** to becoming crazy, **k zvracení** to throwing up.

Syllabic prepositions are generally not vocalised, not even before the same consonant as the one at the end of the preposition, e.g. **od doby** from the time. They are generally vocalised only before the pronouns **všechen** all and **já** I, as well as in a few fixed collocations before certain words beginning with the same consonant as the one at the end of the preposition, e.g. **ode dávna** since long ago, **ode dveří** from the door, **ode dřeva** of wood, **ode dna** from the bottom, **ode dneška** from today, **přede dvěma** before two, **pode trupem** below the trunk, **beze zvuku** without a sound, **beze slova** without a word, **beze studu** without shame, **beze strachu** without fear, **beze starosti** without a worry, **beze změn** without changes. But even in these collocations, with the exception of the pronoun **já**, a certain variation can be observed, e.g. **bez/beze starosti** without a worry, **před/přede dveřmi** right outside the door, **před/přede dvěma** before two. In a very few instances vocalisation occurred in other contexts, e.g. **ode chvíle** since the moment, **beze hnutí** without movement. The preposition **skrze** through does not belong to the same type, as the final **e** is historically a full vowel which was lost at a much later stage.

Kučera (1982) begins by stating that it is a broadly accepted view that the vocalisation of prepositions in Czech and Slovak does not involve any semantic differences between the alternating forms, that it is a purely phonological matter. He states his arguments against this view in Kučera (1984). The prevailing view can be explained by the fact that this problem has not been given due attention by linguists and the phenomenon is only marginally, inexactly, and sometimes incorrectly described in more recent works. The most detailed discussion can be found in Trávníček (1949: 46ff.) Kučera only deals with asyllabic prepositions before a word beginning with a single consonant followed by a vowel. Kučera points out that according to Trávníček the prepositions always vocalise if the following word begins with the same consonant, voiced or voiceless, as the one in the preposition. The prepositions **s** and **z** are also used more often in their vocalised form than in

their non-vocalised form before words beginning with *š-* and *ž-*. If the following word begins with another consonant the prepositions normally do not vocalise, with the exception of *ve městě*, where the two consonants are pronounced [vɛmniɛstɛ] or [vɛmjestɛ]. Trávníček also adds that vocalisation is influenced by the last sound in the word preceding the preposition and that vocalisation is less frequent before a complement than before a noun. Kučera doubts the correctness of the first comment, but agrees with the second.

Kučera then goes on to discuss Havránek and Jedlička (1963: 305), who say that the asyllabic prepositions are vocalised before a word beginning with the same or a similar consonant to the one in the preposition. Kučera points out that they do not define “similar” or give any information about the frequency of the vocalisations. Šmilauer (1972: 33) says, more precisely, that the asyllabic prepositions vocalise before the same (voiced or voiceless) consonant as the one in the prepositions, and that *s*, *z* sometimes also vocalise before *š-*, *ž-*. The same information is given in both SSJČ’60 and SSČ’94. Kučera compares this to standard descriptions of the corresponding vocalisation in Slovak, and finds that the descriptions are basically the same. Nothing more precise is said about the frequency of the vocalisations of *s*, *z* before *š-*, *ž-*. Kučera therefore excerpted samples of these two prepositions followed by words with an initial *š-* or *ž-* followed by a vowel. He found that vocalisation occurred in 57% of the Czech examples and in 98% of the Slovak examples. He mentions a claim by Letz (1950: 50) saying that they should always be vocalised in this context in Slovak. In Czech it can further be observed that vocalisation occurs in 64% of the examples before a noun, but in 43% before a complement. Kučera concludes by claiming that the choice between the forms is influenced by factors such as the speaker’s age, education, origin and how common the word after the preposition is, and most probably certain other factors as well.

Kučera (1984) provides a shorter detailed investigation and discussion of the vocalisation of the two prepositions *s* and *z* before words beginning with *š-*, *ž-* or *ř-* followed by a vowel. He explains this limitation with the fact that this is the only context in which asyllabic prepositions occur in both forms before the same single consonant followed by a vowel, hence it is the simplest context in which variation can be found. He states that asyllabic prepositions never vocalise before a vowel. The two prepositions *v* and *k* exhibit a regular pattern: *v* always vocalises before words beginning with *v-* and *f-*, whereas *k* always vocalises before *k-* and *g-*. Otherwise these two prepositions do not vocalise before a single consonant followed by a vowel. Before a consonant cluster, all asyllabic prepositions display a far more complicated picture, and therefore such contexts have been left for future investigation, after having shed necessary light on these simpler situations.

Kučera argues that the traditional assumption that asyllabic prepositions are vocalised in order to facilitate pronunciation is not sufficient. Based on a questionnaire answered by 52 respondents he gained 5,325 individual

collocations involving the context to be investigated, i.e. the preposition **s** or **z** before a word beginning with **s-**, **z-** (102 relevant answers), **š-**, **ž-** (4,187 relevant answers) or **ř-** (874 relevant answers).

The vocalisation rate before **s-**, **z-** turned out to be only 89%, and before **ř-** as much as 16%, both results in conflict with recommendations in codificational handbooks. No systematic patterns could be detected with regard to the non-vocalised form of the prepositions before **s-**, **z-**. The relatively high rate of vocalisation before **ř-** was mainly due to a couple of fixed phrases: **utržený ze řetězu** (54%) *going on the rampage* and **silnice sjízdná jen se řetězy** (54%) *a road only passable with chains*.

When it comes to the situation before **š-**, **ž-** followed by a vowel, the general rate of vocalisation is exactly 50%. Kučera suggests a few new reasons for vocalisation, besides the traditional facilitation of pronunciation. First he discusses the principle of unambiguity. Taking into account the principles of tone assimilation, the two prepositions **s** and **z** are pronounced identically before **š-** or **ž-**, but if vocalised they are clearly distinguished. He shows that the rate of vocalisation of **z** is 56%, whereas the rate of **s** is only 46%. He argues that the endings in the genitive case, which is governed by **z**, are formally less unambiguous than the endings in the instrumental case, governed by **s**. Therefore the need to disambiguate the preposition **z** is greater and this preposition is vocalised more often. Along the same line he finds that the rate of vocalisation before nouns is higher (52%) than before complements (47%). With a complement there are two case endings giving information about the case, and indirectly about the underlying, possibly assimilated preposition. With no complement there is only the case ending on the noun, which is more often ambiguous than an ending on an adjective. Therefore, the need to disambiguate is greater before a noun than before a complement. The rate of vocalisation before cardinal numbers, which have identical case endings in the two cases discussed, was 65%.

It was further found that vocalisation was more frequent before uncommon words than before common. The rate of vocalisation was 37% before words belonging to the 1,000 most common and 50% before words found in the span of the 9,000-10,000 most common words.

The respondents' education was also shown to influence the choice of form of the prepositions. The nine respondents with only primary education vocalised at a rate of 34%, the 29 respondents with a secondary education at a rate of 43%, and the 14 respondents with a university education at a rate of 52%. This is explained by Kučera as a higher degree of unambiguity in the language that higher education promotes.

Vocalisation is also somewhat higher when the prepositional collocation is found in the stressed position at the beginning of a sentence, namely 55%, as compared to other positions where the rate is 50%.

A small geographical difference was also found. For respondents from

western Bohemia and Prague, the rate was 42%, for eastern Bohemia 45% and Moravia 46%.¹⁰⁰

Differences between age groups were also found, but without any systematic pattern: 14-24 years 41%, 25-34 years 42%, 35-44 years 60%, 45-54 years 43%, 55-70 years 46%.

A few phrases were found where the rate clearly differed from the general rate in the same context: **vajíčka se šunkou** (83%) *ham and eggs*, **nervy ze železa** (87%) *nervs of steel*, **utržený ze šibenice**¹⁰¹ (76%) *going on the rampage (?)*, **nejkrásnější z žen** (0%) *the most beautiful of the women*. These may best be interpreted as partially fixed phrases with lexicalised vocalisation or lack of vocalisation.

It is finally added that no significant differences were found based on the length of the following word or the quality of the vowel following the initial consonant. The author is however convinced that there are differences between spoken and written usage. Therefore other investigations would have to be dedicated to the situation in the spoken language. It is also clear that the usage is extremely unstable and individual, as the respondent with the individual lowest rate of vocalisation used the vocalised form at a rate of only 5% (woman, Nový Jičín, 33 years old, university education) whereas the highest individual rate was 94% (woman, České Budějovice, 51 years old, secondary education).

8.3. Summary of Czech Data

The findings based on the corpus data and the review of the previous literature on the subject will now be summarised in order to give an easily accessible survey of the vocalisations which occur in modern standard Czech.

The lively discussion of this issue in Czech throughout the last century, with many diverging views and suggested principles, makes it easy to understand that the situation has not had any chance to stabilise. A less expected form of a preposition can obviously be used as a stylistic tool with an archaic, poetic, literary, playful, or even somewhat absurd or illiterate effect.

Sixteen prepositions exhibit two alternating forms, with or without an additional final vowel. These prepositions fall into three groups: the syllabic **bez(e)**, **nad(e)**, **od(e)**, **pod(e)**, **před(e)**, **přes(e)**, **zpod(e)**, and the assyllabic **k(e)/ku**, **s(e)**, **v(e)**, **z(e)**, which exhibit systematic patterns of vocalisation, and the five prepositions **poblíž(e)**, **skrz(e)**, **dík(y)**, **krom(ě)**, and **navzdor(y)**, of which four which have not proven to exhibit such systematic patterns. Instead,

¹⁰⁰ One wonders, however, how these figures, as well as some others, are possible, when the general rate of vocalisation in the whole data is 50%, as stated earlier.

¹⁰¹ This seems to be a rather rare expression as it cannot be found in CNK.

the last three almost exclusively occur in the longer shape **díky**, **kromě**, and **navzdory**, whereas **poblíž** is almost exclusively used in its shorter form. Finally, the preposition **skrz(e)** *through* has two main meanings, and the form **skrz** is practically always used in its concrete spatial meaning, whereas the form **skrze** is highly preferred in the somewhat more abstract, instrumental meaning.

The form **ku** is mainly encountered in a function which is best treated as a separate lexeme, namely in expressions of numeric relations of the type **tři ku jedné** *three to one*, where it practically always occurs in this form irrespective of the shape of the following word. Further, the form **ku** also occurs as a vocalised form of the preposition **k**, but only in a handful of lexicalised collocational types, listed at the end of this chapter.

Just as in Polish and Slovak, the grounds for vocalisation fall into two main categories: phonological and lexical. Only the four asyllabic prepositions vocalise on phonological grounds, whereas vocalised forms of both asyllabic and syllabic prepositions occur on lexical grounds. An important difference in comparison to Polish is that, like Slovak, semantic differences between collocations with the vocalised form of the preposition and the corresponding collocations with the non-vocalised form are not so characteristic.

8.3.1. Phonological Vocalisations

The investigation of the corpus data has shown that the conditions for phonologically triggered vocalisation of the asyllabic prepositions **k**, **s**, **v** and **z** in Czech differ considerably from the conditions in Slovak and Polish. Czech, like Slovak and Polish, exhibits certain phonological contexts in which these prepositions practically never vocalise, and other contexts in which they practically always vocalise. But unlike Slovak and Polish, Czech also exhibits quite a large number of phonological contexts in which these prepositions tend to vocalise to a considerable extent. This third type of context is specific to Czech, and is not present in the Slovak and Polish systems. It means that Czech partially exhibits a freer system with instances in which phonology itself is not decisive for the form of the preposition. Instead, in certain contexts, phonology plays the role of establishing an option between the two forms of the prepositions, and it is then rather a question of stylistics and established common usage, but also of individual choices, which one of the permissible forms is actually realised. Here it is important to point out that both the corpus data and the literature on the subject clearly indicate that in these instances, the rate of vocalisation is higher before nouns than before corresponding derivations functioning as complements.

It has also become clear from the data investigated and the literature reviewed that the problematic part of the vocalisation of Czech asyllabic prepositions lies in contexts before words beginning with a consonant cluster.

In contexts before a vowel or a single consonant followed by a vowel or a syllabic sonorant, vocalisation does not pose any major problems. Therefore all initial consonant clusters which triggered vocalisation of the asyllabic prepositions in the previously investigated corpus have been reviewed in the following tables based on data from CNK.

As has already been pointed out, a quantitative investigation of the preposition **s(e)** has not been possible due to the homonymy of the vocalised form **se** with the extremely frequent reflexive pronoun **se**. The available literature supports the view, however, that there is considerable similarity between the behaviours of the two prepositions **s** and **z**.

	ke	ve	ze
tm.*	29 – 7%	10,411 – 65%	1,946 – 61%
dn.*	14,043 – 61%	48,158 – 46%	33,097 – 75%
dr.*	7,494 – 24%	228,769 – 56%	7,374 – 8%
dv.* ¹⁰²	25,756 – 79%	173,033 – 95%	76,834 – 63%
tv.*	743 – 7%	32,728 – 51%	139 – 1%

Table 176. CNK data as of 16 July 2014 regarding vocalisation before clusters of plosive + sonorant.¹⁰³

According to Sawicka (1974: 90), there are many other possible initial combinations of a plosive and a sonorant in Czech words, such as /pl/, /bl/, /tl/, /dl/, /kl/, /gl/, /pr/, /br/, /tr/, /kr/, /gr/, /pn/, /pň/, /tn/, /tň/, /kn/, /kň/, /gn/, /dm/, /km/, /gm/, /kv/, /gv/. Many of them are very frequent, but did not appear in the corpus data. Also, among the high frequencies in the above table, there is not really a general tendency to vocalise on phonological grounds before the consonant clusters in question, but rather, the high frequencies are in most cases due to a few lexicalised collocational types containing the vocalised form of the preposition, and there are many examples of other lexemes which practically never occur with the vocalised form of the preposition. Before /tm/, the vocalisations occur with the root {tm}, mainly in the lexeme **tma** *darkness*; before /dn/ vocalisation is mainly due to the two root {dn} *day, bottom*; before /dr/ it mainly occurs with the lexeme **druhý** *second*, but also, with lower frequencies, some derivations; before /dv/ it mainly occurs with the lexeme

¹⁰² As the second element in a cluster, /v/ has been considered a sonorant as assimilation with respect to voice does not take place. As the first element, however, /v/ assimilates, and has therefore been considered a fricative.

¹⁰³ In order to make the tables more lucid, only the absolute and relative frequency of the collocations with the vocalised form of the preposition is given in this and the following tables.

dva *two*, but also many derivations. Only before /tv/ does there seem to be a connection between the /v/ of the preposition and the identical second element in the initial cluster, even though vocalisation is more frequent before certain lexemes than others, so that vocalisation of **v** is more frequent before the lexemes **tvrzení** *claim*, **tvar** *shape*, and **tvář** *face*.

	ke	ve	ze
cl.*	7 – (13%)	27 – 7%	36 – 21%
čl.*	1,353 – 16%	1,747 – 7%	5,403 – 19%
cv.*	859 – 81%	1,448 – 85%	970 – 67%

Table 177. CNK data as of 16 July 2014 regarding vocalisation before clusters of affricate + sonorant.

There are also a few other possible initial combinations of an affricate and a sonorant: /cm/, /čr/, /čn/, /čň/, /čm/, /čv/; however, none of them seems very frequent.

	ke	ve	ze
hl.*	797 – 3%	316 – 0%	36 – 0%
chl.*	330 – 14%	804 – 5%	32 – 0%
vl.*	514 – 2%		29 – 0%
hm.*	192 – 27%	559 – 10%	22 – 3%
chm.*	27 – (31%)	210 – 47%	6 – 1%
vm.*	16 – (35%)		9 – 7%
hn.*	413 – 12%	182 – 4%	8 – 0%
vn.*	261 – 7%		50 – 1%
hr.*	3,267 – 10%	1,089 – 2%	32,960 – 36%
chr.*	204 – 14%	16 – 1%	1 – 0%
vr.*	139 – 1%		271 – 1%
hv.*	1,875 – 66%	6,329 – 91%	1 – 0%
chv.*	315 – 57%	65,958 – 99%	30 – 2%

Table 178. CNK data as of 16 July 2014 regarding vocalisation before clusters of fricative + sonorant.

The solid squares in the above table show that vocalisation is 100% due to the identity between the preposition and the initial consonant of the following word. In this group of consonant clusters, the frequencies are generally not very high. The major exceptions are the very high rate of vocalisation of the preposition **v** before clusters with the same phoneme as the second element of the cluster following the preposition. There are also rather high frequencies of the vocalised form **ke** before the clusters ⟨hv⟩ and ⟨chv⟩, but this seems to be mainly in the context of the lexemes **hvězda** *star* and **chvála** *praise*.

	ke	ve	ze
jm.*	2,057 – 87%	11,445 – 99%	1,833 – 72%
ln.*	9 – (64%)	96 – 78%	213 – 71%
mn.*	36,093 – 82%	21,404 – 21%	4,756 – 11%
rv.*	892 – 87%	885 – 99%	190 – 71%
lv.*	95 – (96%)	156 – 98%	135 – 83%

Table 179. CNK data as of 16 July 2014 regarding vocalisation before clusters of sonorant + sonorant.

In this group all the frequencies are rather high, with the exception of **v** and **z** before ⟨mn⟩, where the vocalisations are due the pronoun **mně/mě/mne**. Before other lexemes beginning with /mn/, vocalisation is extremely rare.

	ke	ve	ze
lh.*	270 — 71%	3,500 — 97%	169 — 42%
ls.*	78 — (89%)	13 — (76%)	14 — (88%)
ms.*	30 — (71%)	34 — (53%)	683 — 99%
mš.*	102 — 95%	110 — 66%	195 — 84%
mz.*	737 — 91%	1,578 — 88%	3,056 — 99%
lž.*	456 — 96%	552 — 96%	1,913 — 99%
rt.*	651 — 98%	124 — 87%	1,065 — 98%
rč.*	15 — (83%)	44 — (92%)	25 — (71%)

Table 180. CNK data as of 16 July 2014 regarding vocalisation before clusters of sonorant + obstruent.

In this group, the frequencies are generally high, but in the following group they are considerably lower in many of the contexts.

	ke	ve	ze
bř.*	3,014 – 62%	30 – 0%	1,637 – 13%
dř.*	663 – 25%	1,806 – 16%	10,591 – 48%
kř.*		967 – 4%	234 – 2%
tř.*	14,848 – 69%	204,241 – 87%	87,324 – 76%
ps.*	1,117 – 17%	646 – 3%	2,663 – 31%
kš.*		85 – 15%	9 – 7%
pš.*	3 – (8%)	2 – 1%	133 – 16%

Table 181. CNK data as of 16 July 2014 regarding vocalisation before clusters of plosive + fricative.

Again, certain lexemes are responsible for a very large portions of the vocalisations before a certain consonant cluster, e.g. before /tř/, it is the

numerals **tři** *three* and **třetí** *third* in their different case forms, that are responsible for the majority of the vocalisations. Before /bř/ it is the preposition **k** in combination with the lexeme **břeh** *shore* that generates the majority of vocalisations.

	ke	ve	ze
tb.*	0 – –	10 – (27%)	0 – –
kb.*		1,328 – 65%	134 – 30%
kt.*		296,910 – 93%	77,056 – 84%
pt.*	3 – 1%	4 – 1%	16 – 0%
kd.*		2,523 – 69%	1,175 – 49%
tk.*	105 – 76%	424 – 36%	23 – 5%

Table 182. CNK data as of 16 July 2014 regarding vocalisation before clusters of plosive + plosive.

The group in the above table is intriguing, as the variations are quite large, and it seems that high absolute frequency tends to go hand in hand with high relative frequency. The consonant clusters are also rather uncommon, and each type of consonant cluster is only represented by a small number of lexemes, e.g. for /kt/ the lexeme **který** *which* dominates and is only accompanied by a few derivations.

Before moving on to fricative-initial clusters, it should be pointed out that the previous corpus data has shown that the preposition **k** vocalises at a rate of 93% and **v** at a rate of 99% before a sibilant followed by any consonant which is not a syllabic sonorant.

	ke	ve	ze
hř.*	7,336 – 86%	55,702 – 97%	13,865 – 75%
chř.*	75 – 35%	51 – 14%	38 – 6%
vch.*	2,172 – 97%		282 – 59%
vs.*	13,667 – 96%		13,082 – 99%
vš.*	26,005 – 91%		130,837 – 99%
vz.*	22,515 – 77%		29,137 – 99%
vž.*	14 – (39%)		22 – (31%)

Table 183. CNK data as of 16 July 2014 regarding vocalisation before clusters of fricative + fricative.

Once again the preceding and the following group exhibit high absolute and relative frequencies in most cases. There are however also some instances of low absolute and relative frequencies.

	ke	ve	ze
hb.*	20 – (67%)	25 – (50%)	0 – –
vt.*	137 – 53%		48 – 16%
vp.*	138 – 72%		21 – (30%)
vk.*	241 – 72%		35 – 1%

Table 184. CNK data as of 16 July 2014 regarding vocalisation before clusters of fricative + plosive.

	ke	ve	ze
vč.*	2,678 — 66%		6,668 — 89%

Table 185. CNK data as of 16 July 2014 regarding vocalisation before clusters of fricative + affricate.

	ke	ve	ze
džb.*	25 – (69%)	238 – 91%	280 – 94%
ct.*	2,339 – 96%	107 – 64%	268 – 82%
čt.*	13,274 – 93%	389,972 – 100%	73,244 – 97%
čk.*	1 – (100%)	680 – 91%	6 – (100%)

Table 186. CNK data as of 16 July 2014 regarding vocalisation before clusters of affricate + plosive.

Having noticed quite considerable variation before different types of consonant clusters we will move on to have a closer look at the vocalisation of the asyllabic prepositions before root nouns with a root not containing any stable vowel. This type of root noun is repeatedly encountered in the data and in the literature. In order to shed more light on these instances, these nouns will be systematically investigated based on CNK data.

k(e) cti 2,247/1 ¹⁰⁴	s(e) ctí 5,200/0	v(e) cti 30/0	z(e) cti 160/0	<i>honour</i>
k(e) dni 6,499/7	s(e) dnem 697/198	v(e) dni 157/4	z(e) dne 26,825/15	<i>day</i>
k(e) dnu 7,063/3		v(e) dně/dnu 375/0	z(e) dna 5,966/19	<i>bottom</i>
k(e) hře 5,173/28	s(e) hrou 1,594/5,164	v(e) hře 55,115/60	z(e) hry 32,858/144	<i>play, game</i>
k(e) lsti 72/1	s(e) lstí 50/0	v(e) lsti 3/0	z(e) lsti 11/0	<i>trickery</i>

¹⁰⁴ Number of occurrences with the vocalised form to the left and with the non-vocalised form to the right.

k(e) lvu/lvovi 43/1	s(e) lvem 608/7	v(e) lvu/lvovi 3/0	z(e) lva 29/0	<i>lion</i>
k(e) lži 326/0	s(e) lží 269/1	v(e) lži 353/1	z(e) lži 1,390/2	<i>lie</i>
k(e) mstě 25/2	s(e) mstou 7/0	v(e) mstě 10/3	z(e) msty 671/1	<i>revenge</i>
k(e) mši 81/0	s(e) mší 137/5	v(e) mši 33/0	z(e) mše 120/4	<i>mass</i>
k(e) mzdě 473/2	s(e) mzdou 322/12	v(e) mzdě 575/0	z(e) mzdý 2,819/2	<i>salary</i>
k(e) psu 54/8	s(e) psem 9,978/37	v(e) psu 8/4	z(e) psa 392/7	<i>dog</i>
k(e) psovi 59/208		v(e) psovi 4/32		
k(e) švu 10/0	se švem 36/0	ve švu 107/0	ze švu 3/0	<i>seam</i>
k(e) vsi 1,835/1	s(e) vsí 85/1	v(e) vsi 6,325/0	z(e) vsi 1,835/1	<i>village</i>
k(e) vši 2/0	s(e) vší ¹⁰⁵ —	v(e) vši 10/0	z(e) vši 3/1	<i>louce</i>
k(e) zdi 5,717/0	s(e) zdí 366/0	v(e) zdi 4,606/0	z(e) zdi 2,733/0	<i>wall</i>

Table 187. CNK data as of 28 July 2014 regarding vocalisation of asyllabic prepositions before nouns without a stable vowel in their root.

It becomes obvious from the above table that with only a couple of exceptions, the vocalised form of the preposition is generally favoured before this type of root noun. Of course, the figures concerning the number of occurrences with *se* also contain a few occurrences with the reflexive pronoun, but a shorter manual investigation has indicated that such occurrences are very few, thanks to the high rate of co-occurrence between the instrumental case of the noun and this preposition.

The observations made so far can now be summarised into the following general principles.

Phonologically triggered vocalisations of the asyllabic prepositions **PRACTICALLY ALWAYS OCCUR** in the following contexts: 1) before a word beginning with a consonant which is identical to the consonant in the preposition, disregarding the feature of voice, which means that *k* always

¹⁰⁵ Not possible to retrieve as this form of the noun *veš louce* is homonymic with the form *vší* of the pronoun *všechn* *all*.

vocalises before /k/ or /g/, **s** and **z** always vocalise before /s/ and /z/, **v** always vocalises before /v/ and /f/; 2) **v**, **s** and **z** always vocalise before any consonant cluster beginning with a sibilant, i.e. /s/, /z/, /š/ or /ž/; 3) **s** and **z** always vocalise before a /v/ which is followed by one of the three sibilants /s/, /z/ or /š/.

Phonologically triggered vocalisations of the asyllabic prepositions **OCCUR WITH VERY HIGH FREQUENCY** in the following contexts: 1) before root nouns which do not contain any stable vowel (with the exception of **s hrou** *with the game* and **k/v psovi** *to/in the dog*); 2) before clusters of more than two consonants; 3) before less common clusters of two consonants, e.g. /čt/, /rt/, /rv/, /kt/, /jm/; 4) **k** also vocalises before a sibilant followed by another consonant.

Phonologically triggered vocalisations of the asyllabic prepositions are **ALMOST EQUALLY AS FREQUENT** as the non-vocalised form of the preposition in the following contexts: 1) **s** and **z** before /š/ and /ž/ followed by a vowel or a syllabic sonorant; 2) **s** and **z** before /vž/.

Phonologically triggered vocalisations **PRACTICALLY NEVER OCCUR** in the following contexts: 1) before a word beginning with a vowel; 2) before a word beginning with a single consonant followed by a vowel or a syllabic sonorant, disregarding the instances already mentioned above.

8.3.2. Vocalisations by Analogy

In Czech, just as in Slovak, there exist a number of collocations with a high rate of vocalisation, absolute as well as relative, without any phonological grounds for the vocalisation, or, in Czech, with only weaker tendencies to vocalise in the phonological context in question. This means that the rate of vocalisation is low before certain word forms and high before other word forms with the same initial consonant cluster. It seems that for each such context there is one dominating collocational type, which exhibits a very high rate of vocalisation, often close to obligatory. Such collocations must be interpreted as lexicalisations. These collocations most often consist of a preposition and a noun. Alongside with these collocations there are other collocations involving the same roots, that exhibit a lower, but still significant rate of vocalisation. They may be interpreted as vocalisations by analogy with the dominating collocational types. These collocational types more often contain adjectives, diminutives and other derived lexemes.

A good example is the vocalisation of asyllabic prepositions before the root {dvor} *courtyard*, where the vocalisation is very high in the collocational types **ve dvoře** *in the courtyard*, and where different derivations containing the same root morpheme exhibit lower rates of vocalisation.

Further, there seems to be another analogical tendency in Czech, such that phonological contexts may also be subject to analogy and thereby generate

lower frequencies of vocalisation. For the preposition **bez**, it seems that this kind of analogy may well be responsible for the large number of highly individual occurrences of collocational types involving the vocalised form **beze** followed by words beginning with a sibilant-initial consonant cluster, which is also the most frequent initial cluster in the lexicalised collocations with this preposition. The preposition **k** tends to vocalise with a low frequency before words beginning with a consonant cluster with /h/ or /x/ as the first element, e.g. **ke hradu** *towards the castle* with a rate of 20%.

8.3.3. Lexical Vocalisations

All other instances of vocalisation, apart from the above mentioned phonologically triggered ones, must be encoded in the lexicon of modern standard Czech.

The following list constitutes a summary of the information in the literature and the corpus data which have been accounted for earlier in this study. For obvious reasons it can be but a qualified approximation, as the data from the different sources are sometimes somewhat contradictory. Neither does the list claim to give the complete inventory of lexicalisations in Czech. Especially with regard to the preposition **s(e)**, there are with certainty many more collocational types that should be added to the list.

8.3.3.1. Formal Lexicalised Vocalisations

Formal lexicalisations refer to instances of non-phonological vocalisation where the semantics of the collocation does not differ from the sum of the semantics of its constituent parts and where no limitations can be observed with regard to the contexts in which the collocation may be used, i.e. the same constituents do not occur in a non-vocalising collocation with another meaning or in certain other contexts. Of course, the optional formal lexicalised vocalisations occur with the non-vocalised form as well, but this optionality refers to the very same meaning and context.

8.3.3.1.1. Obligatory

The vocalised form is considered obligatory in the following lexicalised collocational types, according to several sources:

- before the pronominal forms **mě/mne/mně/mnou** *me*: **beze mě, beze mne, nade mě, nade mne, nade mnou, ode mě, ode mne, pode mě, pode mne, pode mnou, přede mě, přede mne, přede mnou, zpode mě, zpode mne, ke mně, se mnou, ve mě, ve mne, ve mně, ze mě, ze mne**;

as well as in the following collocations, and also analogically at a lower rate in derivations from the same root:

- **ke dni** *to the day*, **ze dne** *from the day*;
- **ke dnu** *towards the bottom*, **ze dna** *from the bottom*;

- **ke dveřím** *towards the door*, **ve dveřích** *in the door*;
- **ve dvoře** *in the courtyard*;
- **ke hře** *towards the game*, **ve hře** *in the game*;
- **ve hvězdách** *in the stars*;
- **ve chlévě** *in the cowshed*;
- **ve chvíli** *in the moment*;
- **ke jménu** *to the name*, **ve jménu** *in the name*, **ze jména** *from the name*;
- **ve lhůtě** *within the time limit*;
- **ke Lvovu** *towards Lviv*, **ve Lvově** *in Lviv*, **ze Lvova** *from Lviv*;
- **ze lži** *from the lie*;
- **ze msty** *from the revenge*;
- **ze mzdy** *from the salary*;
- **ze rtů** *from the lips*;
- **ve tmě** *in the dark*, **ze tmy** *from the dark*;
- **ke všemu**, **ke vší**, **ke všem** *towards all*.

Czech also exhibits a very small number of lexicalised collocations with obligatory use of the non-vocalised form of the prepositions, despite very strong general tendencies towards the use of the vocalised form:

- **s sebou** *with oneself*;
- **s hrou** *with the game*;
- **k psovi** *to the dog*, **v psovi** *in the dog*.

8.3.3.1.2. Prevailing

Optional, but more frequent than the corresponding collocational type with the non-vocalised form of the preposition, are:

- **beze cti** *without honour*;
- **ode dne** *from the day*;
- **ode dneška** *from today*;
- **ode dna** *from the bottom*;
- **beze jména**, **beze jmen** *without name(s)*;
- **přeše mě**, **přeše mne** *through me*;
- **ve prospěch** *to the benefit*;
- **beze škody** *without damage*;
- **beze švů** *without seams*;

as well as the following collocations, and also analogically at a lower rate in derivations from the same root:

- **ke**, **ze břehu** *towards, from the shore*;
- **ze čtvrt*** *from the fourth*;
- **ve čtvrtek** *on Thursday*, **ze čtvrtka** *from Thursday*;
- **ke čtyřem** *to four*, **ze čtyř** *of four*;
- **ke dvěma** *towards two*, **se dvěma** *with two*; **ve dva**, **ve dvou** *in two*, **ze dvou** *of two*;

- **ve druhém, ve druhé, ve druhých** *in the second*;
- **ve dřevě** *in the wood*, **ze dřeva** *of wood*;
- **ze dveří** *from the door*;
- **ze dvora** *from the courtyard*;
- **ke, ve, ze džbánu** *to, in, from the jug*;
- **ke hřišti** *towards the playground*, **ze hřiště** *from the playground*;
- **ve Kbelích** *in Kbely*, **ve Kdyni** *in Kdyně*, **ve Křtinách** *in Křtiny*;
- **ve městě** *in the town*;
- **ve mlýně** *in the mill*;
- **ke třem** *towards three*, **ve třech** *in three*, **ze tří** *of three*;
- **ve třetím, ve třetí, ve třetích** *in the third*;
- **ve třídě** *in the class*, **ze třídy** *from the class*;
- **ke vchodu** *towards the entrance*.

8.3.3.1.3. Equally Distributed

Optional and equally as frequent as the corresponding collocation with the non-vocalised form of the preposition are:

- **nade dnem** *above the bottom*;
- **nade dveřmi** *above the door (position)*;

as well as the following collocations, and also analogically at a lower rate in derivations from the same root:

- **ke členům** *to the members*;
- **ke druhému, ke druhé, ke druhých** *towards the second*;
- **ke chvále** *to the praise*;
- **ze mlýna** *from the mill*;
- **ke třetímu, třetí, třetím** *towards the third*; **ze třetího, třetí, třetích** *from the third*;
- **ze vrat** *from the gate*.

8.3.3.1.4. Infrequent

Optional and less frequent than the corresponding collocational type with the non-vocalised form of the preposition are:

- **přede dnem** *before the day*;
- **pode dnem** *under the bottom*;
- **nade dveře** *above the door (direction)*, **ode dveří** *from the door*;
- **ke prospěchu** *for the benefit*;
- **se řetězy** *with chains*;
- **beze světla** *without light*;
- **ode všeho, ode vší, ode všech** *from all, from everything*, **nade vš*** *over everything, above everything*;

as well as the following collocations, and also analogically at a lower rate in derivations from the same root:

- **ze čelnů** *from the members*;

- **ze druhého, druhé, druhých** *from the second*
- **ve trojici** *in the trio.*

8.3.3.1.5. Rare

Optional and very infrequent are:

- **přede dvěma** *before two*;
- **přede dveře** *in front of the door* (direction);
- **zpodě dveří** *from beneath the door*;
- **ku lásce** *to love.*

8.3.3.2. Contextual Lexicalised Vocalisations

In this category, most vocalised collocations are semantically motivated, but not all. Some of them have a meaning that does not differ noticeably from the meaning of their constituent parts, but instead the collocation with the vocalised form is only used in certain restricted contexts. It is sometimes also the case that a contextual lexicalisation contains some morphological irregularities which are not present in the corresponding collocation with the non-vocalised form of the prepositions used in the literal meaning of the collocation's constituent parts.

8.3.3.2.1. Obligatory

The following collocational types may be considered practically obligatory:

- **ve dne** *in daytime*;
- **přede dveřmi** *just around the corner* (temporal);
- **ku podivu** *strangely enough, surprisingly enough*;
- **ku příkladu** *for example*;
- **beze slova** *without a word* (manner);
- **beze sporu** *without dispute* (manner);
- **beze všeho** *without hesitation, certainly, of course* (manner);
- **nade vše, nade všechno** *highly, extremely, more than anything* (degree);
- **nade vši pochybnost** *beyond all doubts*;
- **nade vším** *above all*;
- **především** *above all*;
- **beze zbraně, beze zbraní** *unarmed, without weapon(s)* (manner);
- **beze zbytku** *without remainder* (manner);
- **ode zdi ke zdi** *from wall to wall.*

The obligatoriness of these collocations is based either on their very high rate of vocalisation or on implicit recommendations in the literature.

8.3.3.2.2. Frequent

Optional, but prevailing, are:

- **ode dveří ke dveřím** *from door to door*;

- **ku prospěchu** *for the benefit (of something)*;
- **beze spánku** *without any sleep (manner)*;
- **beze smyslu** *without any sense (manner)*;
- **beze spěchu** *without any hurry, calmly (manner)*;
- **beze stopy, beze stop** *without any trace(s) (manner)*;
- **beze strachu** *without fear, bravely (manner)*;
- **beze strát** *without losses (manner)*;
- **beze studu** *without shame, shamelessly (manner)*;
- **beze svědků** *without witnesses (manner)*;
- **přese všechno** *in spite of all (circumstance)*;
- **přede všemi** *in front of everyone*;
- **beze vší + noun** *without any + noun (manner)*;
- **beze změny, beze změn** *without change(s) (manner)*;
- **utrhnout se ze řetězu** *go on the rampage*.

8.3.3.2.3. Infrequent

Optional and less frequent are:

- **přede dveřmi** *in front of the door (spatial meaning)*;
- **být, mít, vzít ku pomoci** *be of help, have, take to help*;
- **ku Praze** *towards Prague (spatial)*;
- **ku příležitosti** *at the occasion (of something)*;
- **beze stínu něčeho** *without even a shade of something*;
- **přese vš*** *in spite of all (circumstantial meaning)*;
- **beze všech + noun** *without any + noun (adverbial of manner)*.

9. Phonological Analysis

In the previous chapters, the contexts in which phonology is involved in the triggering of vocalisation of Polish, Slovak and Czech prepositions were investigated and defined. The aim of this chapter is to discuss these phonological contexts in more detail, i.e. to discuss the phonological principles responsible for these vocalisations, as well as possible reasons for the differences between the three languages.

As my attention for many years has been drawn to the ongoing, highly interesting discussions of different types of vowel-zero alternations couched in the framework of Government Phonology (GP) (e.g. Brockhaus 1995; Charette 1991; Cyran 1994, 2010; Harris 1994; Rowicka 1999; Scheer 1996, 2000; Yoshida 1991), it has been a natural choice for me to use this framework for my investigation. I will however not go into more theoretical detail than is needed, concerning the framework itself, in order to discuss the issues to be investigated. Good general introductions can instead be found in Charette (1991: 5-27), Kaye & Lowenstamm (1981), Kaye et al. (1985, 1989), as well as in *Phonology* (1990, vol. 2), where all the articles in the volume are dedicated to GP.

However, in order to be able to discuss the vocalisation of prepositions in phonological terms, we need some general background on such alternations. I will therefore begin by reflecting on the general nature of phonological representation of vowel-zero alternations. Having established a working representation for the following discussion, the next step will be to discuss the phonological principles which lead to the individual types of vocalisations in prepositions in the three languages investigated.

9.1. The Representation of V-Ø Alternations

The simplest type of vowel-zero alternations may be observed in the Czech masculine noun **den** *day* and the corresponding N.pl. **dny** *days*, or the feminine noun **hra** *play* and the corresponding G.pl. **her** *of the plays*. In order to describe these alternations, there exist three principal strategies: 1) vowel insertion (e.g. Bethin 1992: 123; Rubach 1993: 240ff.), 2) vowel deletion, or 3) surfacing of a vowel which in one way or another is present in the underlying structure, but under certain conditions is “invisible” in the surface structure (e.g. Szpyra 1989: 203; Cyran 2010: 119).

Vowel insertion is problematic, as it is difficult to define under what phonological circumstances such a phonological principle would take action. Rubach’s (ibid.) discussion neatly illustrates this fact. The following instances of N.sg. and G.pl. forms in Czech may serve as simple examples of the fact that insertion is not easily predictable, taking nothing but the phonological context into account.

(31)	<i>no insertion</i>		<i>insertion</i>	
	banka - bank	<i>bank</i>	šunka - šunek	<i>ham</i>
	konzerva - konzerv	<i>can</i>	barva - barev	<i>colour</i>
	sorta - sort	<i>kind, sort</i>	karta - karet	<i>card</i>

As can be seen, in the same consonant clusters insertion applies in some lexemes, but fails to apply in others. In order to handle such difficulties, different principles have been proposed which take into account information such as the origin of the lexeme (domestic or foreign) or domain boundaries before derivational suffixes. For the words in (32), however, it seems impossible to apply even these explanations, as all of the words must rather be considered to be of foreign origin, and none of them can be decomposed as containing derivational suffixes.

There are also differences between the individual languages that are difficult to explain in terms of insertion.

(32)	Czech	Polish	
	firma - firem	firma - firm	<i>company</i>
	karta - karet	karta - kart	<i>card</i>
	barva -barev	barwa - barw	<i>colour</i>

More or less the same types of problems arise if the process is not understood as insertion, but as deletion. This can be illustrated by a few simple examples consisting of Slovak masculine nouns in the N.sg. and G.sg.

(33)	<i>deletion</i>		<i>no deletion</i>	
	bobor - bobra	<i>beaver</i>	tábor - tábora	<i>camp</i>
	meter - metra	<i>metre</i>	náter - náteru	<i>paint</i>

Deletion applies in certain lexemes, but fails to do so in identical contexts in other lexemes.

There are also differences between the individual languages that are just as difficult to explain in terms of deletion as in terms of insertion.

(34)	<i>Polish</i>	<i>Slovak</i>	<i>Czech</i>	
	lew - lwa	lev -leva	lev - lva	<i>lion</i>
		ker - kra	keř - keře	<i>bush</i>

Of course, any of the examples above could just as well be interpreted as exhibiting either insertion or deletion.

Hence, there are many problematic instances which cannot easily be explained in a straightforward manner by mechanisms of insertion or deletion.

I therefore agree with Gussmann (2007: 180) that the only reasonable way to handle the Slavic vowel–zero alternations is to assume that underlying representations contain certain positions where such alternations will occur. As we will see presently, in most cases such positions may be predicted naturally, based on conditions on the syllable structure which are imposed by GP, and only under some conditions is there a systemic possibility of having lexemes with or without such an alternating vocalic position.

Hence, in accordance with GP, I will adopt the standpoint that vowel–zero alternations are instances of realisation versus non-realisation of empty nuclear positions in the underlying syllable structure. Under certain conditions an empty nucleus¹⁰⁶ remains phonetically empty, whereas under other circumstances it is phonetically realised. Let us compare how the two Czech forms **pes** *dog* (N.sg.) and **psa** (A.G.sg.) are represented in GP.

(35)	Cz. pes <i>dog</i> (N.sg.)	vs.	psa <i>dog</i> (A.G.sg.)
	O N O N		O N O N
	x x x x		x x x x
	p e s		p s a

Notice that in the above example, the syllabic structure consisting of onsets and nuclei is the same in both forms, the difference being the position of an empty nucleus.

In these examples the phonological representations are divided into different levels or tiers in order to give a clear picture of all the different parts of the structure, i.e. the syllabic constituents, the duration in time and the phonic material. Such schemes should be interpreted in the following way.

(36)	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;"></td> <td style="text-align: right; vertical-align: top;">Cz. táta [ta:ta] <i>dad</i></td> </tr> <tr> <td style="vertical-align: top;">constituent level (with syllabic constituents)</td> <td style="text-align: center;">O N O N</td> </tr> <tr> <td></td> <td style="text-align: center;"> \ </td> </tr> <tr> <td style="vertical-align: top;">skeletal tier (with timing positions or x-slots)</td> <td style="text-align: center;">x x x x x</td> </tr> <tr> <td></td> <td style="text-align: center;"> / </td> </tr> <tr> <td style="vertical-align: top;">melodic¹⁰⁷ tier (with phonic segments or melody)</td> <td style="text-align: center;">t a t a</td> </tr> </table>		Cz. táta [ta:ta] <i>dad</i>	constituent level (with syllabic constituents)	O N O N		\	skeletal tier (with timing positions or x-slots)	x x x x x		/	melodic ¹⁰⁷ tier (with phonic segments or melody)	t a t a
	Cz. táta [ta:ta] <i>dad</i>												
constituent level (with syllabic constituents)	O N O N												
	\												
skeletal tier (with timing positions or x-slots)	x x x x x												
	/												
melodic ¹⁰⁷ tier (with phonic segments or melody)	t a t a												

The example in (36) shows how the phonic segments or the melody on the

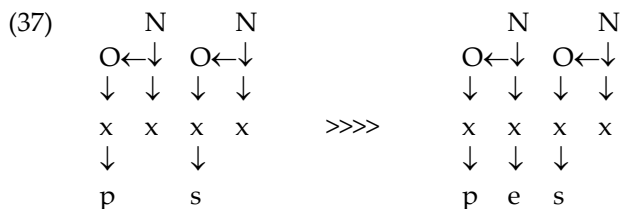
¹⁰⁶ ‘Empty nucleus’ should be understood as short for ‘nuclear skeletal position without attached melody’.

¹⁰⁷ Often also referred to as the segmental tier. This term however seems too vague, as the positions on the skeletal tier could also be regarded as segments.

melodic tier are associated with the timing positions or x-slots, representing duration in time, on the skeletal tier. These x-slots, together with their segments, are in turn associated with the different syllabic units – onsets (O) or nuclei (N) – on the constituent level.

GP operates only with the simple syllabic structure ONSET + NUCELUS.¹⁰⁸ There may be no onset without a nucleus, and practically no nucleus without an onset. The internal structure of these two basic constituents is also restricted, and they may only branch once, thus severely limiting the complexity of phonological structures. Instead, GP operates with empty categories, i.e. both onsets and nuclei may be void of any melodic content associated with the skeletal position. In the present investigation, however, only empty nuclei will be of any interest to the discussion.

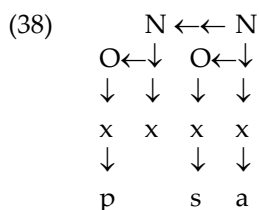
According to GP, there are also far-reaching restrictions applying to the well-formedness of branching onsets and nuclei. Consonant clusters which cannot be interpreted as well-formed branching onsets must be divided by an underlying empty nucleus, i.e. a nucleus lacking melodic content and not realised in the surface structure. Such an empty nucleus in its turn has to be licensed by a following full nucleus, i.e. a nucleus having a melodic content, in order for the empty nucleus to remain phonetically unrealised. This particular type of licensing is called PROPER GOVERNMENT. (Kaye 1990: 138ff.) If, on the other hand, an empty nucleus fails to be properly governed, i.e. it is not followed directly by another nucleus with a realised melodic content, it is forced to realise phonetically. This is exactly what happens in the Czech word **pes** *dog*, where the underlying empty nucleus intervening between [p] and [s] – two consonants which together may not constitute a well-formed onset – is not properly governed, as the following final nucleus is void of any melodic content. The example below shows how the underlying structure /pØ.sØ/ must surface as [pɛs].



In the form **psa**, however, the very same underlying empty nucleus between /p/ and /s/ is properly governed by the final nucleus which contains the melodic content /a/. Thanks to this vowel, the underlying structure /pØ.sa/ is

¹⁰⁸ Traditional codas are not recognised by GP. Instead, a final consonant is interpreted as an onset followed by an empty nucleus. See e.g. Harris & Gussmann (1998).

well-formed and realised as [psa].



Hence, the underlying structure of **pes** is assumed to be the bisyllabic /pØ.sØ/ with two empty nuclei, where the first empty nucleus, which is not properly governed by a following governor, is forced to surface with melodic content. The final empty nucleus, however, may remain phonetically uninterpreted, as final empty nuclei do not need to be licensed by a following nucleus with melodic content. In many languages final empty nuclei are licensed by a language-specific parameter, reflecting the fact that typologically, final empty nuclei occur naturally in some languages, whereas they are less natural or impossible in others.

In most languages, there is a default value for unlicensed empty nuclei: Czech and Polish insert [ɛ] (Gussmann & Kaye 1993: 427), whereas Serbo-Croatian inserts [a] (Diklić 1979: 56), French inserts [ə] (Charette 1990: 235), and Moroccan Arabic inserts [i] (Kaye 1990: 140). The quality of the inserted vowel is regulated by a language-specific parameter.

In Slovak, however, the matter of realisation of vowels in empty nuclei is more complicated, as two different vowels, [ɛ] or [ɔ] (Rubach 1993: 42f.), occur on a regular basis, and, more seldom, even other vowels are encountered. In some instances these vowels also undergo lengthening. A thorough discussion of the mechanisms determining the quality and quantity of the inserted vowel in Slovak is beyond the scope of the present investigation. As pointed out, among others, by Rubach (1993: 137), the prediction of the inserted vowel is a complicated issue, and in accordance with the general principles of GP, I will therefore assume that Slovak does not have empty nuclei positions void of melodic content, as do Czech and Polish, but instead has nuclear positions with a floating melodic content, i.e. a melodic content which is not associated with the x-slot. This means that the melodic material is present in the underlying structure of Slovak, but not attached to the nuclear position, hence as default it is phonetically unrealised.

Summing up, one could argue that the standpoint of GP on the interpretation of vowel-zero alternations is a combination of a syllabic position with no attached phonetic content in the underlying representation, and a mechanism which under certain conditions inserts or attaches phonetic content into such positions. The phonological principles do not determine where such realisations occur, only when they occur. The phonological

framework, however, poses severe restrictions on possible syllabic structures with regard to consonant clusters, therefore in some respects it indirectly determines the necessity of many empty nuclei positions.

In accordance with these general principles of GP, the underlying syllabic representation of the asyllabic prepositions must contain a final empty nucleus, as the underlying structure of any lexical material must contain a nucleus which licenses the presence of the onset. The structure of the prepositions investigated will therefore be as shown below.

(39)Cz. k <i>to, towards</i>	s <i>with</i>	v <i>in</i>	z <i>of, from</i>
N	N	N	N
O←↓	O←↓	O←↓	O←↓
↓ ↓	↓ ↓	↓ ↓	↓ ↓
x x	x x	x x	x x
↓	↓	↓	↓
k	s	v	z
/kØ/	/sØ/	/vØ/	/zØ/

Generally, the empty nucleus of these prepositions may remain phonetically unrealised, as the language-specific parameter in the three languages investigated allows for final empty nuclei. However, as we have already seen in the preceding analysis of the corpus data, the final empty nuclei of the asyllabic prepositions become phonetically realised in a number of contexts. Having established a general phonological model for the representation of vowel–zero alternations, we will proceed to investigate what principles can be assumed to cause the surfacing of a vowel in the final empty nuclear position of the asyllabic prepositions.

9.2. First Vocalisation Principle

The data in the previous chapters clearly show that there is a common major principle regulating a majority of the phonologically driven vowel–zero alternations in asyllabic prepositions in Polish, Slovak and Czech. This principle is applied in contexts where the initial melodic content of the following word is identical or very similar to the melodic content of the preposition. The actual effects of this principle, however, differ between the three languages.

In Slovak and Czech, the asyllabic prepositions always vocalise when the following word begins with an identical consonant or its voiced/voiceless counterpart, whereas in Polish, they only vocalise in this context if such an initial consonant of the following word is also followed by yet another consonant. When the following word begins with a single consonant, the

preposition is instead pronounced as a long consonant together with the initial consonant of the following word. Of course, if applicable, assimilation with regard to voice takes place. Some examples of these instances are presented in (40) below.

(40)	Cz.	Sk.	Pl.	
	ve vodě	vo vode	w wodzie	<i>in the water</i>
	ve filmu	vo filme	w filmie	<i>in the film</i>
	ve vlasech	vo vlasoch	we włosach	<i>in the hair</i>
	ve Francii	vo Francúzsku	we Francji	<i>in France</i>
	se synem	so synom ¹⁰⁹	z synem	<i>with the son</i>
	se ziskem	so ziskom	z zyskiem	<i>with profit</i>
	ze stolu	zo stola	ze stołu	<i>from the table</i>
	ze zlata	zo zlata	ze złota	<i>of gold</i>
	ke komu	ku komu	--	<i>to whom</i>
	ke garáži	ku garáži	--	<i>towards the garage</i>

Before seeking to formulate the phonological principles leading to the above realisations, we will consider the systemically possible realisations that can be expected based on general assumptions within GP. As a point of departure we will take the preposition /v/ followed by a morpheme beginning with /v/, which is in its turn followed by a vowel, as shown in (41). In this context, the domain-final empty nucleus of the preposition is licensed by parameter. As a result, the two identical onsets are adjacent on the melodic tier.

(41)	O	N		O	N	...
			+			
	x	x		x	x	...
	v			v	o	...

According to the so-called Obligatory Contour Principle (OCP) (Leben 1973, Goldsmith 1990: 307ff.), identical phonological material may not appear in adjacent melodic positions. OCP instead leads to the merger of the two neighbouring identical melodies into one single melody associated with two skeletal positions. The simplest possible result of OCP having taken effect on the structure under discussion is shown in (42) below.

¹⁰⁹ In Slovak, underlyingly /zo sinom/, /zo ziskom/, pronounced as [zəsinəm] [zəziskəm].

(42)	O N	+	O N ...	>>>>>	O N O N ...
	x x		x x ...		x x x x ...
					\ /
	v		v o ...		v o ...

The merger of the two identical melodies gives rise to a single melodic content associated with two skeletal positions, a structure which should phonetically be realised as a long consonant.

This is exactly what happens in four of the Polish examples in (40) above. The structure of the first example is shown below in (43).

(43)	Pl. w wodzie [<i>v:ɔɖzɛ</i>] <i>in the water</i>																	
	The underlying structure	(a)	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="width: 20%; text-align: center;">O N</td> <td style="width: 10%; text-align: center;">+</td> <td style="width: 20%; text-align: center;">O N O N</td> </tr> <tr> <td style="text-align: center;"> </td> <td></td> <td style="text-align: center;"> </td> </tr> <tr> <td style="text-align: center;">x x</td> <td></td> <td style="text-align: center;">x x x x</td> </tr> <tr> <td style="text-align: center;"> </td> <td></td> <td style="text-align: center;"> </td> </tr> <tr> <td style="text-align: center;">v</td> <td></td> <td style="text-align: center;">v o dź e</td> </tr> </table>	O N	+	O N O N				x x		x x x x				v		v o dź e
O N	+	O N O N																
x x		x x x x																
v		v o dź e																
	leads to the realisation	(b)	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="width: 20%; text-align: center;">O N O N O N</td> </tr> <tr> <td style="text-align: center;"> </td> </tr> <tr> <td style="text-align: center;">x x x x x x</td> </tr> <tr> <td style="text-align: center;">\ / </td> </tr> <tr> <td style="text-align: center;">v o dź e</td> </tr> </table>	O N O N O N		x x x x x x	\ /	v o dź e										
O N O N O N																		
x x x x x x																		
\ /																		
v o dź e																		

In the surface structure found in (43b), the attachment of the melodic content /v/ to the two x-slots needs to be licensed, as does every part of a phonological structure with the exception of the head of the structure (Harris 1994: 156). In this case, the licensing power must emerge from the following nucleus containing a phonetically realised vowel.

Another possible result of the same underlying structure is a surface realisation in which, for language-specific reasons, the attachment of the same melodic content to two x-slots fails to be licensed. If not licensed, the structure does not comply with well-formedness criteria and must therefore readjust to the principles of the phonological system of the language in question. The melody is then detached from the leftmost x-slot, which is further away from the licenser, which lacks sufficient licensing power. The resulting empty onset and the following empty nucleus together constitute a superfluous syllabic structure, and as such they are deleted as a result of resyllabification of the whole structure. These steps are shown below in (44).

(44)	O N O N ... + x x x x ... v v o ...	>>>>>	O N O N ... x x x x ... \ / v o ...	
	>>>>>	O N O N ...	>>>>>	O N ...
		 x x x x ... v o ...		 x x ... v o ...

This type of surface structure is actually the most natural phonetic realisation in modern standard Slovene (Toporišič 2000: 86) in instances where the preposition *z* *of, with*¹¹⁰ is followed by a word beginning with /z/ or /s/, but for greater clearness a long consonant may also be pronounced. For instance, Slovene **z zobom** *with the tooth* is generally pronounced ['zo:bɔm] as in (45d), but may also be pronounced ['z:o:bɔm] as in (45b), when greater clearness is desired, i.e. in more carefully articulated speech or in order to avoid ambiguity.

(45)	(a)	O N O N ... + \ x x x x x ... / z z o bom
	(b)	O N O N ... \ x x x x x ... \ / / z o bom

¹¹⁰ The two semantically different prepositions have been collapsed in the same way as in Polish and – phonologically but not orthographically – in Slovak. In Slovene the spelling is phonetic, so that *s* is written in both meanings before words beginning with a voiceless consonant.

(c) O N O N ...
 | | | | \
 x x x x x ...
 | | /
 z o bom

(d) O N ...
 | | \
 x x x ...
 | | /
 z o bom

The three languages investigated, however, do not apply this type of surface realisation, which leads to phonetic neutralisation, i.e. loss of the phonetic material representing the preposition. Only the Czech lexicalised collocation **s sebou** [sɛbɔu] *with oneself* might be considered an instance of this type.

The three languages categorically avoid this second possibility and opt for a third possible surface structure, which may also result naturally from general principles of GP, namely the blocking of OCP through the realisation of the intervening empty nucleus, as shown in the Slovak example in (46) below, where the empty nucleus of the preposition is phonetically realised despite the fact that its domain-final position in the preposition would otherwise allow it to be properly licensed and to remain silent.

- (46) Sk. **vo vode** [vɔvɔdʲɛ] *in the water*
 The underlying structure (a) O N O N ...
 | | + | |
 x x x x ...
 | | | |
 v v o dʲɛ
- leads to the realisation (b) O N O N ...
 | | | |
 x x x x ...
 | | | |
 v o v o dʲɛ

This type of blocking of OCP always occurs in Czech and Slovak in the context under discussion. It occurs in Polish as well, but only when the word following the preposition begins with a consonant cluster. This phenomenon will be discussed in more detail below.

To sum up, the following three possible realisation types have been introduced:

(1) OCP takes effect, which leads to the attachment of the melodic content to two x-slots. If this structure is licensed, a phonetically long consonant is realised.

(2) OCP takes effect but the resulting attachment of the melodic content to two x-slots is not licensed, hence simplification follows. The presence of the preposition is not realised phonetically.

(3) OCP is blocked by the phonetic realisation of the empty nucleus, occurring between the two onsets with identical melodies, in order to avoid (2).

The different realisations of the structures under consideration in the three languages emerge from the fact that the choice between these three possibilities is regulated by language-specific parameters, i.e. settings which decide whether a certain general phonological principle is applied or not in a certain language.

One such phonological difference between the three languages concerns consonant length, and in order to explain the observed differences in the vocalisation of prepositions, it is crucial to take into account that Polish exhibits numerous instances of geminate or long consonants, both word-initially and word-internally, e.g. **ssak** [s:ak] *mammal*, **winna** [vin:a] *guilty* (N.f.sg.), whereas Czech and Slovak exhibit no initial long consonants, and Czech exhibits considerably fewer instances of word-internal long consonants (Mašín 1928: 19f.) than do Polish and standard Slovak, e.g. Pl. **cennik** [tɕɛɲ:ik], Sk. **cenník** [tɕɛɲ:i:k], Cz. **ceník** [tɕɛɲ:i:k] *price list*. It should also be pointed out that in colloquial Slovak long consonants regularly become targets of shortening. Stanislav (1953: 274) begins his discussion of long consonants by stating that the popular language in the major part of Slovakia does not recognise long consonants. In the above example, the spelling of Cz. **ceník** follows the pronunciation, but in many other instances the spelling and related word forms clearly suggest two underlying consonantal elements. According to Hůrková (1995: 31), long consonants are only realised in standard Czech pronunciation at evident morpheme boundaries or if there is a risk for ambiguity. Although not stated explicitly by Hůrková, a difference seems to be made between grammatical morphemes on the one hand and semantic morphemes on the other. A typical example of short realisation is the Cz. noun **panna** [pana] *virgin* with the G.pl. **panen** [panɛn] (Hála 1967: 45), where the second /n/ is most likely not perceived as an individual morpheme any longer.

In (47) it is shown how the underlying structure of the G.pl. form **panen**, with two identical consonantal positions divided by an empty nucleus, are also followed by an empty nucleus corresponding to the zero ending of the f.G.pl. The leftmost empty nucleus is not properly licensed to remain silent, as the potential licenser is also a nucleus without phonetic content and therefore lacks the necessary licensing power. Hence, the leftmost empty nucleus must

surface as a phonetically realised vowel, which in Czech is by default /e/.

(47)	O N O N O N x x x x x x p a n n	gives	O N O N O N x x x x x x p a n e n
------	---	-------	---

In the N.sg., however, the final nucleus contains phonetic content as the corresponding ending is [a], i.e. **panna**. Therefore the preceding empty nucleus is properly governed and remains silent.

(48) (a)

O	N	O	N	O	N
x	x	x	x	x	x
p	a	n		n	a

(b)

O	N	O	N	O	N
x	x	x	x	x	x
			\	/	
p	a		n		a

realised in Pl. and Stand. Sk. as [pan:a]

(c)

O	N	O	N	O	N
x	x	x	x	x	x
p	a			n	a

(d)

O	N	O	N
x	x	x	x
p	a	n	a

realised in Cz. and Coll. Sk. as [pana]

The corresponding Polish word is realised as [pan:a] and the prescribed standard Slovak pronunciation is also [pan:a], though shortening may be readily encountered in colloquial speech. This means that (48b) is well-formed in Polish, but instead (48d) is well-formed in Czech. In Slovak, both realisations are well-formed, depending on the register applied.

Such differences between languages are regulated by language specific parameters which determine whether a certain phonological structure is well-

formed or not. In the type of structure under discussion, language-specific parameters allow or disallow the following nucleus to license a directly preceding long consonant attached to two x-slots. In Polish the parameter is set to allow such structures, but in Slovak and Czech, the same parameter is set to avoid such structures and to completely disallow them in an initial position. The exact structure of the long consonants which do occur word-internally in Czech and Slovak is in itself an intriguing issue, which however falls outside the scope of the current investigation.

Another general phonological difference between languages concerns the type of phonetic realisation opted for if the necessary licensing power for the phonetic realisation of a long consonant is not available. The choice must then be made between the second type of realisation, i.e. shortening of the long consonant, and the third type, blocking of the merger into a long consonant. Some languages, such as Swedish and Slovene, are more inclined to allow morphemes to merge phonetically, whereas other languages, such as Norwegian and English, seem to prefer to preserve each morpheme in the phonetic realisation by blocking certain OCP effects. To see this, one can compare the realisations of the Scandinavian present tense morpheme /r/ after a verb stem ending in /r/: e.g. /çø:r/-/r/ *drives* is realised in Swedish as **kör** [çø:r], as opposed to the corresponding Norwegian realisation **kjører** [çø:rər]. Similarly the genitive /maks/-/s/ is realised in Swedish simply as **Max** [maks] *Max's*, as opposed to the corresponding English realisation **Max's** [mæksɪz]. The same also applies to the English plural morpheme /z/, as well as to the past tense morpheme /d/ in English, in contrast with the Swedish /de/: e.g. English /mend/-/d/ gives **mended** [mændɪd], whereas Swedish /tänd/-/de/ gives **tände** [tɛn:dɛ] *lit.* Such examples are similar to the Slavic /s sestr.../ *with the sister*, which in Czech is realised as **se sestrou** ['sɛsɛstrɔʊ], but in everyday standard Slovene is realised as **s sestro** ['sæ:stɾɔ]. Such differences must be understood as general differences in morpho and phonotactics between individual languages, and the choice between the two possibilities is regulated by language-specific parameters.

Summing up, in Polish a melody attached to two x-slots may be licensed if directly followed by a nucleus with a realised vowel, because a long consonant needs the full licensing power of a following nucleus. Hence the licensing power may not be shared with the dependent constituent of a branching onset. If a melody, which due to OCP becomes attached to two x-slots, is then at the same time part of a branching onset, the following nucleus cannot provide enough licensing power for two complex structures. In example (49) below, N₂ cannot at the same time license both the branching onset O₂ and the melody attached to the two x-slots dominated by O₁ and O₂. Licensing power from O₂ would have to be provided both to the preceding O₁ and to the branching constituent to the right. A general assumption within GP is that licensing cannot simultaneously work in two directions (Harris 1994: 156). To resolve

this conflict, the last of the three solutions presented above is chosen, i.e. to block OCP by phonetically realising the empty nucleus between the two identical melodies.

- (49) (a) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄
 | | | \ | | | | |
 x x + x x x x x x x
 | | | | | | |
 v v w o s a x not well-formed; >>>>>
- (b) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄
 | | | \ | | | | |
 x x x x x x x x x x
 \ / | | | | |
 v w o s a x not well-formed; >>>>>
- (c) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄
 | | | \ | | | | |
 x x x x x x x x x x
 | | | | | | |
 v w o s a x not well-formed; >>>>>
- (d) O₂ N₂ O₃ N₃ O₄ N₄
 | \ | | | | |
 x x x x x x x
 | | | | | | |
 v w o s a x morpheme loss; >>>>
- (e) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄
 | | | \ | | | | |
 x x x x x x x x x x
 | | | | | | | |
 v e v w o s a x OCP blocked by vowel insertion

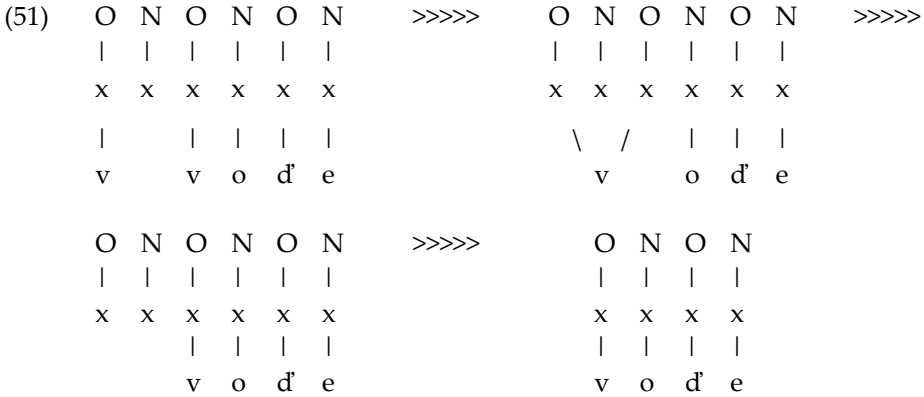
If instead a melody attached to two x-slots is followed by an empty nucleus, this empty nucleus cannot function as the licensor of the long consonant. Again, the application of OCP would lead to shortening of the initial consonant attached to the two x-slots by detachment from the first of the x-slots, and the preposition would become phonetically unrealised.

- (50) (a) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄ O₅ N₅
 | | | | | | | | | |
 x x + x x x x x x x x
 | | | | | | | |
 v f t o r k
- (b) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄ O₅ N₅
 | | | | | | | | | |
 x x x x x x x x x x
 \ / | | | |
 f t o r k not well-formed, >>>>
- (c) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄ O₅ N₅
 | | | | | | | | | |
 x x x x x x x x x x
 | | | | | |
 f t o r e k not well-formed, >>>>
- (d) O₂ N₂ O₃ N₃ O₄ N₄ O₅ N₅
 | | | | | | | |
 x x x x x x x x
 | | | | | |
 f t o r e k semantic loss; >>>>
- (e) O₁ N₁ O₂ N₂ O₃ N₃ O₄ N₄ O₅ N₅
 | | | | | | | | | |
 x x x x x x x x x x
 | | | | | | | |
 v e f t o r e k OCP blocked by vowel insertion

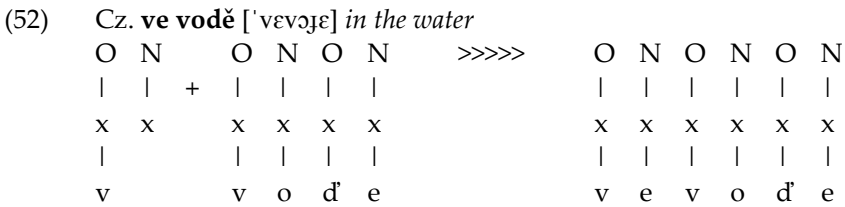
Hence, Polish allows for initial consonant length, but only if the melody representing the long consonant is attached to two non-branching onsets directly followed by a nucleus with melodic content, which can function as the licenser for the long consonant. Therefore, if the asyllabic preposition occurs before a following branching onset or an onset which is followed by an empty nucleus, the preposition is saved from phonetic loss through the realisation of a default vowel in the position of the empty nucleus.

Czech and Slovak, however, are generally less in favour of geminates, and word-initially they are simply not permissible. In phonological terms, this fact must be explained by language-specific parameters not supplying enough licensing power to license a melodic content attached to two subsequent onsets if they are not also preceded by a nucleus with melodic content. If the empty nucleus of the asyllabic preposition will not be realised, OCP forces the

melodic content of the preposition to merge with the identical melodic content of the following onset. Such a structure could not be licensed due to language-specific parameters in Czech and Slovak. The melodic content of the long initial consonant would obligatorily be detached from the first onset. Resyllabification would follow and the initial O + N would be deleted, as they are both void of any content, as shown below.



The risk of phonetically losing the preposition if the final empty nucleus of the preposition remains unrealised is the most plausible reason as to why the empty nucleus is realised in the above context, even though it is licensed to remain silent. A principle that prevents the loss of the semantic content of the preposition overrides the purely phonological principle which would lead to OCP and subsequently to a simplified skeletal structure. Hence, in Czech and Slovak, OCP is prevented from taking action by the realisation of the empty final nucleus of the asyllabic prepositions, as is shown below.



9.2.1. Assimilation of Tone

Assimilation of tone in obstruent consonants is obligatory and regressive in all three languages and there are no phonological grounds for differences in the realisation of the final empty nucleus in the asyllabic prepositions, dependent upon the feature VOICE in an initial consonant following the preposition. This feature automatically spreads from a following initial consonant to the preceding consonant in the preposition. If vocalisation does not take place,

OCP has to take effect. Therefore, in Czech and Slovak, **k** vocalises before /k/ as well as /g/, **v** vocalises before /v/ as well as /f/, and **s** and **z** vocalise before /s/ as well as /z/. For Polish, which allows for initial long consonants, the syllabic prepositions vocalise in the same contexts, but only if the initial consonant of the following word is followed by another consonant, i.e. **w** vocalises before both /v/ and /f/ followed by a consonant, whereas **z** vocalises before both /z/ and /s/ followed by a consonant.

9.2.2. Assimilation of Place of Articulation

Another type of assimilation which may take place in the languages investigated is assimilation with regard to place of articulation of sibilants. Here again, the three languages do not behave in the same way. In Czech, /s/ or /z/ followed by /š/ or /ž/ should not be assimilated with regard to place of articulation. Šmilauer (1972: 240) states that assimilation of sibilants differing in place of articulation, in instances such as **sžíravý**, **rozžehnout**, is non-orthoepic. Nor, according to the Slovak norm, should such sibilants be allowed to assimilate; however, immediately after stating this principle, Pauliny et al. (1968: 43) add that nonetheless this kind of assimilation applies generally in Slovak dialects. In Polish, both types of pronunciation, with or without assimilation before /š/, /ž/, as well as before the typically Polish /ś/ and /ź/, has equal codificational status (e.g. Sawicka 1995: 151; Wiśniewski 2001: 117).

In this context, i.e. when the prepositions **s** and **z** occur before words beginning with /š/ or /ž/, and in Polish also with /ś/ or /ź/, the rate of vocalisation is practically the same as before /s/ and /z/ in both Polish and Slovak. Only in Czech can a significant difference be observed, and this is only so if a sibilant with a different place of articulation is followed by a vowel. In such contexts, the rate of vocalisation is approximately 60%, whereas it is practically 100% if the sibilant is followed by a consonant. The following examples illustrate the situation described.

(53)	Pl.	Sk.	Cz.	
	z żabą	so žabou	s(e) žábou	<i>with a frog</i>
	ze szkoły	zo školy	ze školy	<i>from school</i>

When an initial sibilant with a different place of articulation is followed by a vowel, there is no need for the preposition to vocalise in Polish. If the preposition is assimilated, the emerging long consonant is licensed to be pronounced, as Polish allows for initial long consonants. If the preposition is not assimilated, no problem arises. In Slovak, the urge to assimilate seems to be so high that this threat is systematically prevented by the realisation of the empty final nucleus of the preposition. In Czech, the urge to vocalise is not as strong as in Slovak. The preposition may very well be pronounced without assimilation and then no problem arises. There may also be an urge to

assimilate the preposition, in which case the final empty nucleus of the preposition is realised in order to prevent such an assimilation, and the subsequent shortening of the long consonant, which would mean a phonetic loss of the preposition. The two possible realisations in Czech, [ˈzʒa:bɔɯ] and [ˈsɛʒa:bɔɯ], also lead to two different pronunciations with regard to the place of stress. This difference in itself may of course influence the choice between the forms, as the non-vocalisation of the empty nucleus in the preposition offers the natural possibility of stressing the noun itself and not the preceding preposition. In Slovak, however, the stress is not fixed to the preposition in the same absolute way as in Czech, meaning that in Slovak there is a freer choice between placement of the stress on the preposition or on the noun, i.e. [ˈzɔʒabɔɯ] or [zɔʒabɔɯ], even if the preposition is vocalised in both instances.

The question, however, is why the situation in Czech and Polish looks different if the sibilant is followed by another consonant. It might be that the urge to assimilate becomes reinforced in that type of context because the preposition would need a certain amount of licensing power in order to resist the assimilation, power that is not available because the following vowel's power is used to license the two initial consonants of the word following the preposition. The examples discussed by Sawicka and Wiśniewski, where according to the Polish norm both assimilation and lack of assimilation are equally permissible, only involve contexts where the second of the sibilants is followed by a vowel. Contexts where the second sibilant is followed by another consonant are not discussed, hence it is not clear if lack of assimilation is also permissible in that context.

9.2.4. OCP and Syllabic Prepositions

The effect of OCP does not constitute a severe problem with regard to the syllabic prepositions as they consist of enough phonetic material to be interpretable even if their final consonant is lost phonetically.

9.3. Second Vocalisation Principle

Another phonological context exhibiting a high general frequency of vocalisation of the asyllabic prepositions is discussed in some of the literature. It is also rather evident in the corpus data. This context occurs when the consonant phoneme of the asyllabic preposition also occurs as the second element of the initial consonant cluster in the word following the preposition, and when the first element of that cluster is one of the sibilants /s/, /z/, /š/, /ž/, /ś/, /ź/ or the labiodental fricative /v/. The approximate rate of vocalisation in these contexts in the three languages investigated is shown below.

Polish	Slovak	Czech	before
ze 74%	so/zo 99%	ze 99%	/v/+SV
ze 98%			/v/+SC
we 0%	vo 95%	ve 99%	S+/v/
—	ku 50%	ke 100%	S+/k/

Table 188. Rate of vocalisation of syllabic prepositions before certain consonant clusters involving sibilants.

It seems that some mechanism is at work which strives to avoid the repetition of similar melodic content on both sides of certain consonants. The reasons for such tendencies could well be certain general restrictions on the phonotactics of the different languages. That these tendencies are least noticeable in Polish is well in line with Sawicka's (1974: 100ff.) claim that Polish generally exhibits a greater complexity of consonant clusters than do Czech and Slovak.

In the context where the prepositions *s/z* are followed by a word-initial cluster consisting of /v/ followed by a sibilant, all three languages tend to vocalise to a very high degree. Czech and Slovak vocalise practically always, whereas the Polish system once again is sensitive to whether a vowel or a consonant follows after the context described above.

The first question is why vocalisation should be sensitive to this context at all. It is evident that the initial /v/ must have some very special status, as this is actually the only context, apart from the first vocalisation principle discussed above, that leads to an apparently phonologically conditioned vocalisation in Polish. It also seems probable that this special status could be shared by the initial sibilants in the other two contexts in Slovak and Czech, but not in Polish, as shown in Table 188. We will return to this issue shortly.

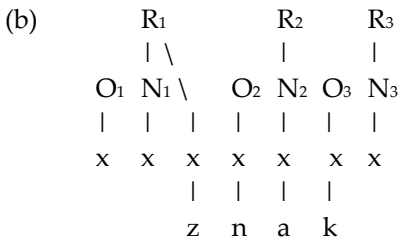
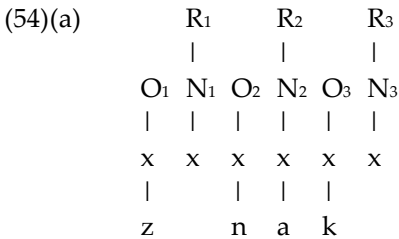
To begin with, similar problems with strange behaviour in initial consonant clusters occur in several languages. Kaye (1992) investigates, among other things, the Italian masculine definite article, which generally has the form *il*, but in certain phonological contexts changes its form to *lo*. This form occurs before consonant clusters containing a sibilant plus another consonant, as well as in some other instances. Clusters of a sibilant plus another consonant are disallowed in many languages and they are avoided by insertion of a default vowel into an empty nuclear position before such consonant clusters in foreign borrowings and names, e.g. **Scotland** is **İskoçya** in Turkish, **Escocia** in Spanish and [eskotland] in Persian. Kaye (ibid.) argues in his article for a special status for some initial, strangely behaving consonants, namely as codas, which in GP are rather called rhymal adjuncts. Such word-initial rhymal adjuncts must form part of a syllable containing both an empty onset and an empty nucleus.

Gussmann (2007: 201 ff.) discusses this issue in Polish and concludes that it is difficult to find any evidence for the underlying syllabification of instances

like the initial cluster of the word **znak** *sign*. From general assumptions within GP, it follows that /zn/ may not be a well-formed branching onset, as there are severe restrictions imposed on any branching constituent. The head and the complement of a branching constituent always find themselves in a governing relationship where the head is the governor and the complement the governee. In order for the governor to govern the governee, it must be stronger than the governee. (Harris 1990)

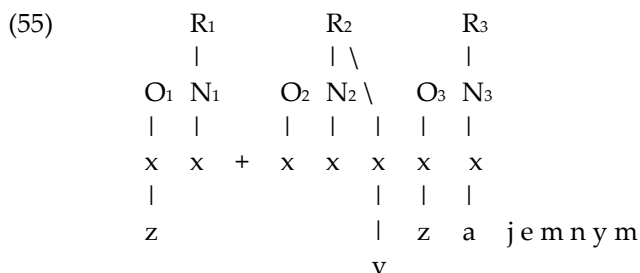
Strength in GP is interpreted as melodic complexity, meaning positions containing melodic content consisting of many phonological elements. Elements are the non-binary parallel to the traditional binary phonological features. In GP, a negative feature in principle equals the absence of the corresponding phonological element. For example, [-voice] is instead understood as the absence of the element voice. Fricatives lack the element [occlusion], whereas both fricatives and stops have the element [noise]. In GP, the traditional sonority scale is understood as a strength scale where the complexity or number of elements of a phonological segment defines its place on the strength scale.

Because sibilants are generally assumed in GP to be among the weakest segments, and /n/ contains both occlusion and nasality, /s/ is not a probable governor for /n/. Therefore /zn/ is ruled out as a branching onset. It could instead be two subsequent onsets with an intervening empty nucleus which is properly governed by the following vowel /a/, as shown below in (54a), but it could also, as in the Italian nouns triggering the special definite article **lo**, be a rhymal adjunct followed by an onset, as shown in (54b).



In GP, onset segments and rhymal segments are also assumed to reside on different tiers, which means that the onset before the rhymal adjunct and the

onset following it are “visible” to each other on their common onset tier. If this is applied to the type of context under discussion for triggering vocalisation of the preposition *z* before a /v/ followed by a sibilant, as for example before the word **wzajemnym** *mutual*, the constituent structure will look as follows.



As the rhyml adjunct /v/ resides on a different tier from O₁ and O₃, the two instances of /z/ may interact with each other even across the intervening /v/. In this particular context, when the consonant cluster is followed by a vowel, two different realisations of the underlying structure are possible, **ze wzajemnym** or **z wzajemnym**, hence the question still remains why the two identical sibilants that are neighbouring on the onset tiers are practically always separated if the second one is followed by a consonant, but only to some extent if it is followed by a vowel. This seems in some way similar to the instances within the first vocalisation principle, where Polish accepts a long initial consonant followed by a vowel, but not followed by a consonant. The exact details concerning the reasons for these two possibilities and for the actual distribution between the two is too complex a question, however, to be investigated in detail at this time. It seems that the type of sibilant in O₃ probably plays an important role. There is also a risk that certain lexicalised collocations influence the proportions of the distribution to an extent that makes it difficult to recognise the truly phonological vocalisations from the lexicalised ones.

Finally, it should be pointed out that the structure of the word **wzajemny** cannot be /vØ.za.je.mØ.ny.mØ/, with an empty nucleus intervening between the two first consonants, as there is no possibility of such a structure triggering vocalisation of the preceding preposition. Asyllabic prepositions are not dependent on the first nucleus of the following word to have any melodic content in order to license the final empty nucleus of the preposition. There is an abundant number of such collocations with the non-vocalised form of the preposition in both Polish and Slovak. There are, however, tangibly fewer such collocations in Czech, a fact which we will come back to shortly. Anyhow, in a structure where /z/ occupies an onset position, it is impossible for the consonant in the preposition to “see” and interact with the consonant in O₃, and there is no reason for the final empty nucleus in the preposition to vocalise, as Polish allows for non-vocalised asyllabic prepositions before

extremely complex consonant clusters, e.g. **w mgle** in *the fog*.

It must also be pointed out that even if the /z/ in O₃ in (55) is visible to the /z/ in O₁, it does not seem to mean that the /v/ in R₂ is invisible to that /z/ in O₁, as we find the expected realisation according to the first vocalisation principle in examples like **we wzajemnym** in *the mutual*, where the /v/ in the preposition obviously must be able to interact with the /v/ in the rhymal adjunct in order to avoid a long /v:/, which in turn would not be licensed due to lack of licensing power, as the following vowel has to license both the /z/ of its onset and the /v/ in the preceding rhymal adjunct.

We will now continue to the preposition **w/v**, which in Polish practically never vocalises before a sibilant followed by a /v/, whereas in Czech it practically always vocalises in the same context, and in Slovak it vocalises at a rate of approximately 95%. The same phonological structure is proposed in these instances, i.e. that Czech and Slovak most likely syllabify these sibilants as rhymal adjuncts, whereas Polish syllabifies them as separate onsets, leading to the different realisations of the final empty nucleus of the preposition.

The Slovak collocation **vo svojom** in *one's own* will serve as an example.

(56)

	R ₁			R ₂				R ₃	
				\					
	O ₁	N ₁		O ₂	N ₂ \		O ₃	N ₃	
	x	x	+	x	x	x	x	x	
	v						v	o	j o m
						s			

Just as in the preceding example, the consonant in the rhymal adjunct resides on a different rhymal tier, and the two onsets, both containing a /v/, are able to interact on the onset tier, where the neighbouring identical onsets are divided by the realisation of the empty final onset of the preposition. There does not seem to be any other reasonable way of explaining this type of vocalisation, but it remains an important question why the vocalisation is not obligatory and whether or not the cluster being followed by a vowel or a consonant might play any minor role, just as it does in Polish.

In the same type of collocation, it seems that the Polish structure is rather a branching onset, as in (57a), and the cluster may be considered to exhibit a progressive assimilation emerging from the /s/ and affecting the /v/. It could however equally well be that /s/ and /v/ are two consecutive onsets, as in (57b), because the preposition's final empty nucleus is not sensitive to the lack of melodic content in the first nucleus of the following word. Instead, the domain boundary after the preposition seems to be enough for the empty nucleus to be licensed just by rights of its final position.

(57)(a)

	R ₁			R ₂	R ₃	R ₄				
O ₁	N ₁		O ₂	N ₂	O ₃	N ₃	O ₄	N ₄		
			\							
x	x	+	x	x	x	x	x	x	x	
v			s	f	o		i	m		gives ['fsfɔim]

(b)

	R ₁			R ₂	R ₃	R ₄	R ₅				
O ₁	N ₁		O ₂	N ₂	O ₃	N ₃	O ₄	N ₄	O ₅	N ₅	
x	x	+	x	x	x	x	x	x	x	x	
v			s		f	o		i	m		gives ['fsfɔim]

The same analysis as in (56) also holds for the corresponding Czech contexts, where the rate of vocalisation is practically 100%.

Finally, we will continue to the context of the preposition **k** before words beginning with a sibilant followed by a /k/. The analysis would be of the same type as given above in (56). In Czech, this preposition always vocalises in this context, whereas in Slovak the rate of vocalisation is only about 50%. This is an intriguing difference that needs further investigation, but it might be suspected that an awareness of the observed general “overusage” of the vocalised form of this preposition in contexts other than in the proclaimed correct contexts, i.e. before /k/ and /g/, would lead to a certain amount of hypercorrection and avoidance of vocalisation in the context under discussion, as the preposition is directly followed by a sibilant and not by a /k/ or a /g/.

Finally, it should be mentioned that there are a few other instances resembling the types of contexts discussed above which also exhibit significant rates of vocalisation in Czech. Such instances are **ve tv***, which exhibits vocalisation at a rate of 30%, and **ve dv*** at a rate of 92%. At the same time, **ze dv*** is vocalised at a rate of 58%, which cannot be explained by the second vocalisation principle. In these instances, it does not seem to be vocalisation on phonological grounds, but vocalisation on lexical grounds, as only collocations with a very small number of lexemes exhibit very high rates of vocalisation. Such lexemes include **dva** *two*, **tvůj** *your*, **tvar** *form*, **tvář** *face* and derivations of the same. Many other less frequent lexemes do not exhibit any particularly high rate of vocalisation. Also, from the point of view of GP, /dv/ and /tv/ are well formed onsets, as the first plosive segment is without doubt stronger than the second non-plosive segment, meaning that /t/ and /d/ contain the element

[occlusion].

Next, there are the two contexts **ve chv***, which exhibits vocalisation at a rate of 99%, and **ve hv*** which has a rate of 87%. Again, in these contexts, the number of lexemes involved is rather low, but it might very well be that the analysis proposed as the second vocalisation principle can and should be applied. However, in order to make such a claim, a detailed study of larger amounts of corpus data involving these contexts would be needed, as well as further theoretical consideration of the arguments in favour of treating /x/ and /h/ as possible rhymal adjuncts followed by an onset containing /v/.

Finally, the contexts **ke sch***, with a rate of 98%, and **ke sh***, with a rate of 91%, look very similar to the contexts discussed above; we will see shortly, however, that they also fit well into the context of the third vocalisation principle, to which we will now proceed.

9.4. Third Vocalisation Principle

Polish and Slovak prepositions do not vocalise in any other phonological contexts than those that can be handled by the first and the second vocalisation principles. Czech prepositions, however, also seem to vocalise in a systematic way in other phonological contexts.

Let us first compare the use of the other asyllabic prepositions in Slovak and Czech in the contexts discussed, in connection with the second vocalisation principle. We will however leave out contexts that are evidently taken care of by the first vocalisation principle.

Slovak	Czech	before
so/zo 99%	ze 99%	/v/+S
ku 16% ¹¹¹	ke 85%	
vo 95%	ve 99%	S+/v/
ku 8%	ke 93%	
vo 1%	ve 99%	S+/k/
ku 50%	ke 100%	

Table 189. Rate of vocalisation of asyllabic prepositions before certain consonant clusters involving sibilants.

In Slovak, as expected, no significant rates of vocalisation occur where the

¹¹¹ This figure is based on the assumption that the collocational type **ku všet*** *to all* should not be taken into account here, as its exceptionally high rate of vocalisation at 76% seems to indicate that this collocational type constitutes an instance of lexicalised vocalisation. If however these collocations are also included, the general rate of vocalisation is 36%.

consonant in the preposition differs from the second consonant of the following initial consonant cluster. The vocalisation rate for **ku** is in line with the general tendency of this preposition to vocalise at a low rate in any context. Let us compare a few examples.

(58)	Slovak		Czech		
	vo svojom	97% ¹¹²	ve svém	100%	<i>in one's own</i>
	ku svojmu	6%	ke svému	93%	<i>towards one's own</i>
	ku škole	68%	ke škole	100%	<i>towards school</i>
	vo škole	0%	ve škole	100%	<i>in school</i>

In Czech the rates of vocalisation are very high if the preposition differs from the second consonant in the following initial cluster. Instead there seems to be something about the very structure of these clusters that triggers vocalisation, i.e. that the mere status of the initial sibilant as a rhymal adjunct would trigger the vocalisation of the preceding preposition, regardless of the quality of the consonantal segments on both sides of the rhymal adjunct.

Let us also consider two other types of contexts that trigger high rates of vocalisation in Czech. One such context is that of nouns which consist of a root without any stable vowel in it. The vocalisation of prepositions before such nouns was investigated in **Fel! Hittar inte referenskälla..** Those data show that in the absolute majority of such collocations there is an overwhelming preference for the vocalised form of the preposition. Summarising all the data in that table gives a general relative rate of 97% in favour of vocalisation before the singular forms of those nouns. Let us also compare a few corresponding examples from Slovak with the Czech ones.

(59)	Slovak		Czech		
	vo hre	0%	ve hře	100%	<i>in the play, game</i>
	vo hrách	0%	ve hrách	70%	<i>in the plays, games</i>
	zo dna	19%	ze dna	100%	<i>from the bottom</i>
	zo dňa	97%	ze dne	100%	<i>from the day</i>

It becomes obvious that in some instances there are considerable differences between the two languages, but it is quite difficult to say anything about the actual reasons for the vocalisations in the individual collocations, because with these highly frequent lexemes there has already proven to be a high probability that collocations with a vocalised preposition may be lexicalised. Looking for general tendencies, however, there is a far more general

¹¹² Frequencies in this and the following examples are based on data from SNK and CNK on 7 August 2014. The relative rate of the given form of the preposition is compared to the other form of the same preposition before the same word form.

vocalisation taking place before this type of words in Czech. This more general pattern of highly frequent vocalisations may well be due to the syllabic structure of this type of words. The alternative would basically be that all the prepositional collocations with words in this category have been lexicalised. In Slovak the situation is clearly different. The rate of vocalisation varies from 0% to almost 100%, hence a common factor in these words as a group cannot be the cause, and individual collocational types within this group, such as *zo dňa from the day*, will have to be treated as lexicalisations in Slovak.

Yet another type of context triggering many vocalisations in Czech is that of words beginning with rather unusual consonant clusters, such as those investigated earlier in section 8.3.1. Before many such clusters, the Czech asyllabic prepositions tend to vocalise to a significant extent, whereas in Slovak almost no vocalisations take place in this type of contexts. Let us compare a few examples.

(60)	Slovak		Czech		
	vo ktorom	0%	ve kterém	95%	<i>in which</i>
	ku včerejšiemu	16%	ke včerejšímu	67%	<i>to yesterday's</i>
	vo džbáne	2%	ve džbánu	88%	<i>in the jug</i>

Such initial clusters triggering a high rate of vocalisation are, from a GP point of view, evidently impossible branching onsets, which means that they have to constitute two consecutive onsets with an intervening empty nucleus. Once again, it therefore seems that Czech asyllabic prepositions may be sensitive to the syllabic structure of the following word.

Judging from the three types of vocalisations discussed in this section, it seems that the Czech asyllabic prepositions, unlike their Polish and Slovak counterparts, are sensitive to the constituent structure of the following word to a significant extent. This would mean that in Czech the final empty nuclei of the asyllabic prepositions are not automatically treated as licensed to remain silent, but instead they need to be properly governed by a following full nucleus. This dependency on the first nucleus of the following word implies that the morpheme boundary after the preposition has a different status in Czech than in Slovak and Polish. In Czech the empty nucleus of an asyllabic preposition is not automatically licensed as domain-final, but in Slovak and Polish it is. This new assumption, however, does not always hold for Czech, as there are often two possible realisations of a certain prepositional phrase involving the kind of initial consonant cluster under discussion.

It has already been shown that vocalisation patterns in Czech are more unstable than in the other two languages, as Czech has been undergoing some far-reaching changes in this respect over the last couple of centuries. It can be suspected that even the status of the domain boundary after the asyllabic prepositions has become subject to a reinterpretation. This may be compared with the stress pattern in modern Czech, where the stress is always placed on

the first nucleus with melodic content, i.e. always on the vocalised asyllabic prepositions, just as on the syllabic prepositions. This means that there is a tighter link between the preposition and the following word in Czech than in Slovak and Polish, which have different stress patterns than Czech. While Czech consequently moves stress to the preposition, Slovak may, but does not necessarily do so, and Polish only does so in a handful of lexicalised collocations.

If Czech is still in some respects undergoing these changes, it might well be that there are two competing systems, an older one which looks more like the Slovak system, where the empty nuclei of the prepositions may still be licensed by parameter as domain-final, and a newer system where this privilege is no longer supported by a domain boundary. This new situation would mean that the Czech asyllabic prepositions have become fully sensitive to the syllabic structure of the following word and that the empty nucleus of the preposition needs to be properly governed by a following nucleus with melodic content.

As an effect of such ongoing changes and two competing systems, it is not strange that the situation with respect to vocalisation of the asyllabic prepositions is unstable. Also, in many other parts of the linguistic system, Czech offers various competing forms, and such differences often play a noticeable role in the rich Czech system of stylistic registers. Czech offers many possibilities of making formal stylistic distinctions through grammatical details, and the vocalisation of asyllabic prepositions can in some phonological contexts be used to achieve such effects.

In those instances where vocalisation and non-vocalisation come to compete in the same phonological context, as in the examples in (60) above, it seems from the previous investigation of the corpus data that certain patterns can be observed. The higher the absolute frequency of a certain collocation, the higher the relative rate of vocalisations seems to be. There have been many examples of collocations with a highly frequent noun with a high rate of vocalisation, accompanied by other derived nouns and adjectives with lower absolute frequencies and also lower relative rates of vocalisation. It therefore seems that in the choice between the two competing systems, the newer one has a stronger position in contexts where it has been able to establish itself through the high absolute frequency of the collocation, whereas in collocational types with a lower absolute frequency, the older system has a stronger influence, because the vocalisations do not feel sufficiently well-established. Hence there seems to be a strong connection between absolute and relative frequency in instances where the application or not of the proposed third vocalisation principle would lead to different competing realisations.

It may also be that there are different competing interpretations of the syllabic structure of a certain word. For example, the adjective **hrací** *playing*

may naturally be interpreted as beginning with a well-formed branching onset /hr/, but from a morphological point of view, there is also evidence in the G.pl. form **her** *plays* that there exists an empty nuclear position between the two consonants in this root. It seems that an interpretation with two onsets is dominant in the noun itself, but that an interpretation with a branching onset is much more natural in the adjective, leading to the preferred forms **ve hře** *in the game*, but **v hracím prostoru** *in the game area*.

Another factor that may play an important role in the choice between the two competing realisations is the fact that for focusing purposes, sentence stress may fall on the word following the preposition. This seems to happen more often with a complement than with a noun, and might therefore be a contributing reason as to why vocalisation is less frequent before adjectives and other complements than before nouns.

Summing up, the proposed third vocalisation principle would thus mean that the asyllabic prepositions are incorporated into the domain of the following word. This principle is only applied in Czech, and it seems to be an innovation from the point of view of the traditional Czech system described in the literature, as well as from the perspective of the other languages investigated. This principle has not yet stabilised, and is not always applied. It is more often applied in frequent collocations, less so in infrequent collocations. The effects of this principle are also considerably influenced by different possible interpretations of the syllabic structure of the initial consonant cluster in the word following the preposition; hence the preposition only vocalises if both the newer system including all three vocalisation principles is applied, and if the following consonant cluster is interpreted as having an empty nucleus after the initial consonant or before an initial rhymal adjunct. At the same time, this means in practice that vocalisation is potentially possible before any cluster of two consonants, as it is not ruled out that there could be an empty nucleus after the first consonant. Analogy might play an important role in the choice of one syllabic structure over the other. Examples of two different possible syllabic interpretations of the same preposition and noun are given in (61) below.

(61)(a)	R ₁		R ₂	R ₃	R ₄			
	O ₁	N ₁	O ₂	N ₂	O ₃	N ₃	O ₄	N ₄
			\					
	x	x	+	x	x	x	x	x
	k		h	r	a	n	i	c
			h	r	a	n	i	c

(b)		R ₁		R ₂		R ₃		R ₄		R ₅	
	O ₁	N ₁		O ₂	N ₂	O ₃	N ₃	O ₄	N ₄	O ₅	N ₅
	x	x	+	x	x	x	x	x	x	x	x
	k			h		r	a	n	i	c	i

With the syllabic structure in (61a), the realisation will be **k hranici** *towards the border*, regardless of whether the third vocalisation principle is applied; with the syllabic structure in (61b), however, the application of the third principle leads to **ke hranici**, whereas the non-application of the same principle leads to **k hranici**. In this collocational type the non-vocalised form is used at a rate of 94%, and it seems that the inclination to treat /hr/ as a branching onset prevails heavily in Czech, unless there is very tangible evidence for another structure, as in the noun **hra** *game* with its G.pl. form **her**, which evidently manifests the empty nucleus between the two onsets.

9.5. Questions for Further Investigation

Having investigated an extensive amount of corpus data and defined the majority of contexts in which the vocalisation of prepositions occurs, as well as having proposed some working hypotheses about possible explanations for the phonologically driven vocalisations, it would now be necessary to continue these investigations in two main directions.

First of all there is a need to investigate the general syllabic structure and phonotactics of Czech, the status of the branching onsets and the strength hierarchy of Czech consonants, in order to better understand the interplay between the vocalisation of the empty nucleus in asyllabic prepositions, the role of the domain boundary after prepositions, and the initial syllabic structure of the triggering word.

At the same time, a detailed investigation of larger amounts of corpus data is needed, concerning some of the more conspicuous contexts in all three languages, where the vocalisation leads to results that are difficult to interpret because there seem to be more subtle factors involved, such as the segment directly following the two initial consonants after the preposition. There is also a need for a thorough corpus investigation of the Czech preposition **s(e)**, as well as a deeper look into the apparent connection between absolute and relative frequency in instances where both forms of a preposition are systemically permissible in Czech.

10. Conclusions

The present investigation has led to the following major observations.

The Czech **poblíž(e)**, **dík(y)**, **krom(ě)** and **navzdor(y)** as well as the Slovak **navzdor(y)** exhibit two alternating forms in free distribution, differing however in stylistic value and frequency.

Only the Czech preposition **skrz(e)** *through* exhibits a slight semantic difference between the two forms.

The majority of the prepositions consist of syllabic prepositions which only vocalise in collocations with certain words. Such collocations are lexicalised and not predictable on phonological grounds.

A smaller number of the prepositions are asyllabic. They have one form without any vowel (Pl. **w**, **z**; Sk. and Cz. **k**, **s**, **v**, **z**) and another vocalised form with a vowel (Pl. **we**, **ze**; Sk. **ku**, **so**, **vo**, **zo**; Cz. **ke/ku**, **se**, **ve**, **ze**).

As well as the preposition **z(e)** *from, of*, Polish also has an adverbial particle **z(e)** *about, approximately*, which vocalises under different circumstances than the preposition, and Slovak exhibits a similar particle which always has the form **zo**.

The four (two in Polish) asyllabic prepositions are most often vocalised on phonological grounds, but to a smaller extent they also vocalise, just like the syllabic prepositions, in collocations with individual word forms, and such collocations must be listed in the lexicon.

The contexts in which the four asyllabic prepositions do vocalise on phonological grounds may be formulated in terms of three vocalisation principles.

The first vocalisation principle blocks OCP, i.e. it blocks the assimilation and merger of the preposition with the initial consonant of the following word into a long consonant in contexts where a long consonant would not be permissible based on the general phonotactics of the individual language. This principle is applied more frequently in Slovak and Czech than in Polish, as Slovak and Czech do not tolerate long initial consonants at all, whereas Polish does tolerate them under certain circumstances.

The second principle is in some respects similar to the first. It prevents segments which are too similar from residing next to each other on the onset tier, despite an intervening rhymal adjunct, by realising an intervening empty nucleus. This principle is also applied more frequently in Slovak and Czech than in Polish, as Polish seems to have fewer initial rhymal adjuncts.

The third principle is unique to Czech and it demands that the asyllabic prepositions should be treated together with the following word so that the form of the asyllabic preposition becomes fully dependent on the syllabic structure of the initial consonant cluster of the following word, instead of

being automatically realised as non-vocalised by default as in Polish and Slovak. According to the third principle the asyllabic preposition can only remain non-vocalised before a single consonant or a cluster that constitutes a well-formed onset. Otherwise it must vocalise. This principle is understood as an innovation in Czech phonology, i.e. there exist two competing systems, an older system without the third principle and a newer system containing the third principle. This choice between two competing systems in Czech leads to a number of possible choices between two equally possible realisations of an asyllabic preposition. In some instances the two possible realisations can be observed to be used with different stylistic effects. It has also been noted that the vocalised form is often more frequent before a basic noun and less frequent before related derived nouns and adjectives. At the same time it seems that a high absolute frequency of a certain collocation with a possible vocalisation also promotes a high relative frequency of that vocalisation. These two tendencies often go hand in hand.

Corpus Texts

Polish Corpus Texts

Technical Regulations

<http://ec.europa.eu/enterprise/tris/>

2007-0276-PL, 2007-0287-PL, 2007-0337-PL, 2007-0341-PL, 2007-0463-PL, 2007-0487-PL, 2007-0603-pl, 200500440.PL, 200500630.PL, 200500632.PL, 200500634.PL, 200501359.PL, 200501372.PL, 200501376.PL, 200501451.PL, 200600251.PL, 200600297.PL, 200600398.PL, 200600402.PL, 200600406.PL, 200600464.PL, 200600488.PL, 200600511.PL, 200600546.PL, 200600852.PL, 200600962.PL, 200600990.PL, 200601281.PL, 200601781.PL, 200601783.PL, 200601847.PL, 200602148.PL, 200602305.PL, 200602307.PL, 200602309.PL, 200602345.PL, 200602347.PL, 200602349.PL, 200602351.PL, 200602403.PL, 200602405.PL, 200602778.PL, 200602826.PL, 200700073.PL, 200700076.PL, 200700264.PL, 200700266.PL, 200700601.PL, 200700700.PL, 200700807.PL, 200700809.PL, 200701013.PL, 200701414.PL, 200701431.PL, 200701954.PL, 200702097.PL, 200702217.PL, 200702219.PL, 200702242.PL, 200702322.PL, 200702589.PL, PL040317, PL041577, PL050075, PL050120, PL050121, PL050122, PL060043, PL060082, PL060084, PL060098, PL060104, PL060108, PL060117, PL060187, PL060269, PL060349, PL060461, PL060462, PL060482, PL060512, PL060513, PL060514, PL060526, PL060527, PL060528, PL060529, PL060596, PL060603, PL070003, PL070046, PL070047, PL070126, PL070229, PL070233, PL070234, PL070235, PL070264, PL071019, PL071021, PL071023, PL071141, PL071149, PL071151, PL071175

Polish Factual texts

Abratowicz, Gabriela. 2010. Nienawidzę poniedziałków.

Brande, Dorothea. 2010. Będę pisarzem.

Fuentes, J.D. 2006. Seksulany klucz do kobiecych emocji.

Gartner, Malwina. 2009. Daj jej orgazm.

Gawryluk, Barbara. 2008. Odchudzanie. Fakty i mity.

Grabka, Anna. 2006. Wykorzystaj potęgę artykułów.

Kijak, Marcin. 2009. Jak uczyć się do egzaminów.

Kijak, Marcin. 2009. Stres.

Kizinczuk, Sergiusz. 2006. Wprowadzenie do magii.

Królak, Paweł. 2006. Sekty. Co warto o nich wiedzieć.

Mart, Piotr. 2008. Kobieta a mężczyzna.

Mikołajewska, Barbara. 1999. Zjawisko wspólnoty.

Mikołajewska, Barbara. 2003. Na początku.

Niwiński, Tadeusz. 2010. Bóg Einsteina.

Odważnie o seksie. 2008.

Ostrowski, Grzegorz. 2005. Gotowanie jest proste.

Szabelski, Marcin. 2009. Sztuka uwodzenia kobiet w internecie.

Szafranowska, Katarzyna. 2006. Sekrety czytania.

Szyborska, Bożena. 2006. Jak rozstawać się z papierosami.

Tomczyk, Ireneusz. 2006. Tylko ty decydujesz.

Urban, Tomasz. 2008. RSS dla blogerów.

Wimmer, Paweł. 2006. OpenOffice org Math.

Polish Fiction

Barczynski, Adam – Slepý los;
Długosz, Artur – Fortel; Ulica;
Dziewulski, Ryszard – Bez pamięci;
Karcz, Łukasz – Rozkład normalny; Banda; Działal obrońca Zajry; Opowieść o najgorszym;
Mroźek, Sławomir – Tango;
Pacek, Stefan – Człowieczeństwo; Mój osobisty wszechświat; Szukanie siebie i świata;
Piskulak, Andrzej – Kluczem do baśni może wytrych; Po trzeciej zmianie; Z gwiazdozbioru oczu;
Dąb jak chłop; Świętokrzyski człowiek; Lirtania; Na ludzkiej ziemi; Trans – Gombrowicz; W imię ojca;
Podolska, Hanka – Jeden dzień w Tokio; Niepewność;
Różewicz, Tadeusz – Kartoteka;
Schulz, Bruno – Sklepy cynamonowe;
Siwek, Katarzyna – Cara Bosca; Moja mała Flo;
Sosińska, Grażyna – Dwie damy na pozór Fosse’a i de Laclosa;
Świech, Andrzej – Tajemnica Marleny;
Wiśniewski, Grzegorz – Obca opowieść wigilijna; Pies czyli kot; Tygrys.

Parliamentary transcripts

11.01.2006 — 20.09.2006

Polish Newspapers

Tygodnik Podhalański: 20.08.1995 — 22.09.1997
Tygodnik Powszechny: 1994, no 1–52.

Slovak Corpus Texts

Technical Regulations

<http://ec.europa.eu/enterprise/tris/>

2006-0162-SK, 2006-0186-SK, 2006-0455-sk, 2007-0001-SK, 2007-0050-SK, 2007-0063-SK, 2007-0140-SK, 200500976.SK, 200501331.SK, 200502748.SK, 200502915.SK, 200502917.SK, 200502919.SK, 200503004.SK, 200503007.SK, 200503009.SK, 200503011.SK, 200503036.SK, 200600818.SK, 200600945.SK, 200601532.SK, 200700063.SK, 200700284.SK, 200700336.SK, 200700667.SK

Slovak Factual texts

Babilon, Ján (?-1877) Prvá kuchárska kniha v slovenskej reči – Diel I; Prvá kuchárska kniha v slovenskej reči – Diel II; Prvá kuchárska kniha v slovenskej reči – úvod a slovníček;
Bujnáč, Pavol (1882–1933) Dr. Karol Kuzmány. Život a dielo; Pavol Országh-Hviezdoslav. K 70-tym jeho narodeninám; Zo slovenskej estetiky;
Hálek, Ivan (1872–1945) Kysuce; Zápisky lekára;
Lajčiak, Ján (1875–1918) Slovensko a kultúra;
Rovnianek, Peter Vítazoslav (1867–1933) Zápisky zaživa pochovaného;
Škarvan, Albert (1869–1926) Zápisky vojenského lekára;
Štefánik, Milan Rastislav (1880–1919) Ekvádorský zápisník;

Vlček, Jaroslav (1860—1930) Dejiny literatúry slovenskej; Kapitoly zo slovenskej literatúry; Listy z Čiech; Literatura na Slovensku, její vznik, rozvoj, význam a úspechy; O Slovensku a slovenskej literatúre; Pavol Jozef Šafárik; Za Janom Nerudom.

Slovak Fiction

- Banšell, Koloman (1850—1887) Atalanta; Šuhaj ako z iskry; Literárnoteoretické štúdie a polemiky; Na parolodi a železnici; Okyptenec; Túhy mladosti; Z tých krajších časov
- Bielek, Anton (1857—1911) Keď sa starí začnú hašteriť; Na mylných cestách; Na salaši; Obecné počty; Obrázky z hôr; Poviestka z hôr; Pre cudzie viny; Rozprávka z Poľany; Z dôb utrpenia
- Bodický, Michal (1852—1935) Rozpomienky a pamäti
- Bodický, Samuel (1850—1919) Nihilizmus; Nová kométa; Opatrná matka; Otcova kliatba; Prvá ľúbosť; Vo väzení; Zaľúbenec; Zlý duch; Zmrŕvych; Žena dvoch mužov
- Braxatoris-Sládkovičov, Martin (1863 — 1934) Zo života malých
- Cambel, Samo (1883—1935) Básne; Moje otcovstvo; Pavlus; Pochybenie; Prekazená služba; Slečna Nina; Stará matka; Trest; V snahe; Vytriezvenie; Z tmavej zápače; Záhorcovo šťastie; Zlomená duša
- Cirbusová, Jolana (1884—1940) Bez roboty; Kožušníci; Len odišiel a nevrátil sa; Najkrajšia, najmúdrejšia, najbohatšia; Pred večierkami; Prispôsobenie; Tulákové radosti; V tieni smrti; Žobráčik
- Czambel, Samuel (1856—1909) Slovenské ľudové rozprávky
- Čajak, Ján (1863—1944) Cholera; Chudobný boháč; Ecce homo!; Fuksi; Ja si môj mliečnik nenačnem; Jeden deň zo života vlasteneckého kňaza; Jedna korunka; Jožkova svadba; Len pekne; Malí zbehovia; Pokuta za hriech; Pred oltárom; Predaj hory; Pytačky; Rodina Rovesných; Strýc Miško; Špitál; Tridsaťpäť dekagramov; Ujčekov Mikušov posledný deň; Únos. Obrázok z dolnozemskeho života; V druhej triede; V tretej triede. Črty; Vohľady; Vtáčie hniezdo; Z povinnosti; Zo života malého nezbedníka; Zuza
- Červeň, Ján (1919—1942) Divé vtáča; Dva snečné dni; Modrá katedrála; Prorok; Sedem listov; Svätá žiara; Z denníka; Zlomený kruh
- Dilong, Rudolf (1905—1986) Budúci ľudia; Cesty vyhnanca; Čierne balady; Gorazd; Helena nosí ľaliu; Honolulu, pieseň labute; Hviezdy a smútok; Ja svätý František; Mladý svadobník; Mladosť z očistca; Moja krv; Nevolať nevolaj; Nezomrie, kto miluje; Opřcha lístie jesenné; Osamelý šiel som svojou cestou; Pieseň lásky; Plač; Pod krížom; Pred nocou labuť spieva; Pred tebou váza; Rozjímania; Ruža Dagmar; S úsmevom Vianoc; Spod Južného kríža; Staroba radí. Epigramy; Vidíte zrnko; Vo výškach života; Vo vetre myšlienky; Zakliata mladosť; Život má múdrosť. Aforizmy
- Hollý, Jozef (1879—1912) Amerikán; Černová; Emancipácia; Geľo Sebechlebský; Kubo; Márnratný syn
- Hranko, Martin (1899—1940) Furko a Murko; Kukučka
- Ivanková, Elena (1871—1941) Barko Mári; Boj so šarkanom; Citróny; Červená blúza; Dcéra pána Karabína; Dobrá partia; Dvom pánom slúžiť; Glosy; Keby nebola vojna...; Kleopatrin peniaz; Koniec; Lúčenie; Menuet; Moderná bájka; Na mena; „Počiatok románu“; Posledná idyla. Obrázok z doby rokokoa; Pred sto rokmi; Prostá historia; Svadobná noc; Svedok lásky; Štedrý večer; Taký je život...; Trio...; Tuberóza; V krinolíne; Všetichsvätých na vojne; Žlto-brokátový plášť
- Jégé, Ladislav Nádaši (1866—1940) Adam Šangala; Aké sú naše cesty?; Ako sa dievčatá vydávajú; Ako sa mladí ľudia ženía; Alina Orságová; Annalena; Babylonská veža; Biela labuť; Boj o pokrok; Brok; Bubuš, Bubulík a Bubulínko; Cesta životom; Cholera; Česť; Články; Daromná robota; Dáždnik; Dáma v hoteli Bellevue; Deľba; Dedinský notár; Dlhá chvíľa; Dobrý deň; Dr. Garvič; Druhou triedou; Duch môjho strýka; Dve vdovy; Eh, Katka!; Gemma docci; Honba; Horymír; Hriešne hnevy; Hriešni ľudia; Jaríkovský kostol; Jedovaté kvety; Kačenka; Kapitalista; Katka volí; Kúra; Kozinský mlyn; Krásne časy rokoka; Kritiky; Krpčky sv. Floriána; Kto mal

pravdu?; Kuruci; Laicizovaný úrad; Láska a hrach; Láska — trápenie; Láska v manželstve; Magda Marošová; Magister rytier Donč; Magnézia; Manželstvo na próbu; Mariella Lucardi; Martin Skalák; Matúš Zelenec; Mía; Mišo žiarli; Mudrci a klinec; Myslíte, že to nie je nebezpečné?; Na jednom brde tkaní; Na operáciu s truhlou; Náušnice; Nero; Niet ľudom pomoci?; Nová veda; Nové topánky; O sebe; Obchod; Obchody; Obyčajné tragikum; Obyčajný príbeh; Odmena; Okolo pravdy; Omyl; Ondro Kubas; Orchidea; Otravné lásky; Otrgajst; Panské šaty; Pavo rozpráva; Pán biskup a východ slnca; Pán svojej vôle; Páni sa zabávajú; Pomsta; Potrestanie Štefana Kršku; Poučenie skúseného človeka; Prečin a výprask; Priateľská návšteva; Prvý bozk; Radosti života; Rozjímanie o snoch; Rozličné osudy; Rozum na cestách; S duchom času; Sami si robíme starosti; Sedliaci a páni; Sestry; Slabí ľudia — mocná láska; Slnce svieti, ako chce; Slovenské aristokratky; Soňa; Spravodlivosť; Stále len matka; Stolica sa smeje; Strieborná lyžička; Svätopluk; Tajný plán; Trionfo della morte; Trnovec; Učím sa nakupovať; V čom je radosť života; Vada; Všetelijaké rozumy; Víkend; Výhľad z Brezovca; Výhody spoločenského života; Vdova a dieťa; Večera; Vianoce Adama Boronču; Wieniawského legenda; Z listov; Záhada; Zásady pani Gagačkovej; Zlá voľba; Zverolekár; Žart

Jesenský, Janko (1874—1945) Čierne dni; Básne; Básne z rokov 1913–1914 a z vojenského obdobia pred zajatím; Cestou k slobode; Demokrati; Dodatky k zbierkam Po búrkach a Proti noci; Dodatky ku zbierke Zo zajatia; Jesenný kvet; Juvenília; Malomestské rozprávky; Náš hrdina; Novely; Ostatná krátka próza; Po búrkach; Proti noci; Verše; Z rokov 1928—1942; Zo starých časov; Zo zajatia

Klein-Tesnoskalský, Belo (1853—1941) Obrazy z dejín prešovských evanjelikov

Korbačka, Jozef (1907—1943) Tri vrecia zemiakov

Kovalik Ústiansky, Ján (1861—1950) Ballady a povesti; Čarodejná moc; Dlhé roky!; Kainovo dielo; Krb; Memento; Z pút k slobode

Krivoš, Vladimír (1865—1942) Tolstojec

Kukučín, Martin (1860—1928) Ako sa kopú poklady; Bacúchovie dvor; Baldo & Comp.; Bohumil Valizlosť; Tábor; Cestopisné črty. Rukopisy; Čas tratiť — čas platiť; Dedina v noci; Dedinský jarmok; Dedinský román; Deti; Dies irae; Do školy; Dojmy z Francúzska. Črty z ciest; Dom v stráni; Dosiaľ nepublikované práce; Drotárove Vianoce; Dve cesty; Hajtman; Hody; Keď báčik z Chochoľova umrie; Keď bol Matej malý; Klbká; Košútky; Komasácia; Koncepty nepublikovaných rukopisov; Koniec a začiatok; Lukáš Blahosej Krasoň; Mať volá; Máje; Mišo; Mladé letá; Na ľade; Na hradskej ceste; Na jarmok!; Na obecnom salaši; Na Ondreja; Na podkonickom bále; Na prielohu; Na stanici; Na svitaní; Neprebudeny; „Nie milí, nie drahí“; O Michale; Obecné trampoty; Obeta; Od trištvrté na osem do ôsmej; Panský hájnik; Parník; Pán majster Obšival; Po šiestej hodine; Po deviatich rokoch; Pod vládou ženy; Pozor na čížmy; Praha v znamení výstavy; Prečo Adam Chvojka spáva teraz už doma; Prechádzka po Patagónii; Prechádzka po výstavisku; Prechádzka po Vlkolíne; Prechádzky po výstavke v Prahe; Pred pekný domec. (Neúplný koncept); Pred skúškou; Preháňanky; Prvá zvada; Punta Arenas; Regrúti; Rijeka — Rohič — Záhřeb; Rodina; Rohy; Rozmajrínový mládnik; Rysavá jalovica; Slepá kura a zrno; Spod školského prachu; Susedia; Svadba; Sviatočné dumy; Syn výtečníka; Štedrý deň; Tela; Tichá voda; Tiene i svetlo; Tri roje cez deň; Úvod k vakáciám; V Dalmácii a na Čiernej Hore; Výlety; Veľkou lyžicou; Vianočné oblátky; Visitatio canonica; Z lovcových zápisov; Z našej hradskej; Z teplého hniezda; Za ženou; Zakáša — darmo je!; Zádruha; Zápisky zo smutného domu; Zo študentských časov; Zo stupňa na stupeň

Maróthy-Šoltéssová, Elena (1855—1939) Články; Moje deti; Na dedine; Popoluška; Prípravy na svadbu; Proti prúdu; Sedemdesiat rokov života; V čiernickej škole; Za letného večera

Maršall-Petrovský, Gustáv (1862—1916) Ako Mišo Záhoran prestal piť; Ako sa v P.....i divadlo hralo; Baronica; Bratranec; Doma; Fi donč!; Jánošík, kapitán horských chlapcov; Kaplán; „Kto mečom bojuje, mečom zahynie!“; „Mí nó spik engliš!“; Obete čiernych diamantov; Ochranca komárov; Okradnutí zlodeji; Papagáj; Pán Tlčhuba v nesnádzach; Pod hrubou kôrou; Podvod; Pozde. Poviedočka zo života amerických Slovákov; Prestál próby; Rýchlikom; Samaritán; Späť; Stratená varta; „Tak sa časy menia...“; Tri dni z denníka; „Ty moje serce...“; V cudzine; Z

- dediny; Z parku
- Milkin, Tichomír (1864–1920) Básnické listy; Clivé tóny; Literárne štúdie a kritiky; Piesne a spevy; Pomník zásluhám; Romance, balady, legendy; V úzkosti o národ; Zakázané sny
- Mitrovský, Milan Thomka (1875 – 1943) Cyntia; Intonácie jari; Lipka; M. Th. M; Maiestas artis. Umenie, veda a literatúra Pani Helène; Sopra Minerva; Z listov bratislavským felibrom; Zo zápisníkov
- Oliva, Pavol Ušák (1914–1941) Čierne kvietie; Oblaky; Tak umieral básnik Paľo Oliva
- Ondruš, Ján (1932–2000) Korenie; Posunok s svetom; Prvý mesiac; V stave žľče
- Poľský, Dlhomír (1864–1935) Memento mori; Piesne národa; Proti prúdu; Quodlibet; Tempa passati; Vesna života; Z rodinného kruhu; Zpod Eadorhory; Jozef Podhradský; Švárna Skaličanka; Holuby a Šulek; Vyhodený študent pri skale
- Prídavok, Anton (1904–1945) Do videnia; Kľúč od srdca; Korešpondencia; Lámané drieky; Podobizne; Pribina; Publicistika; Rozbité ministerstvo; Svitanie na východe; Svitá; Valaška a dukáty; Vrásky času
- Rázus, Martin (1888–1937) Bača Putera; Krčmársky kráľ; Maroško; Maroško študuje
- Roháček, Ján (1869–1939) Cesta do Egypta a Palestíny
- Roy, Vladimír (1885–1936) Výber z básní nepojatých do zbierok, napísaných po prevrate
- Royová, Kristína (1860–1936) Abigajil Karmelská; Ako Kvapôčka putovala; Ako prišli lastovičky domov; Ako zomieral Sláviček; Bez Boha na svete; Deti hauzírerov; Divné milosrdenstvo; Keď nikde nebolo pomoci; Kráľovná zo Sáby; Lótova žena; Navrátený raj; Náman Sýrsky; Opilcovo dieťa; Peterko; Prišiel domov; Splnená túžba; Staniša; Šťastlivé Vianoce; Šťastní ľudia; Šťastný Štedrý deň; To, čo večne trvá; Traja kamaráti; Tuláci; V snehovej krajine
- Socháň, Pavol (1862–1941) Honorár; Svätajanské ohne na Slovensku
- Stacho, Ján (1936–1995) Apokryfy; Dvojramenné čisté telo; Iné básne; Prvé básne; Svadobná cesta; Z prežitého dňa; Zážehy
- Tajovský, Jozef Gregor (1874–1940) ... Aby odvykali; Až do smrti; Apoliena; Čo ho zabilo?; Blúznivci; Bojko; Do kúpeľa; Do konca; Domov!; Dvanásť duší; Dve školy; Hnevník; Horký chlieb; Hrdina; Hriech; Jano Mráz; Jej budúci; Jej prvý román; Korteš; Kosce Mócik; „Kšo, kšo!“ alebo: „Kto má škodu, má aj posmech“; Lipky na hrobe Sama Chalupku; List; Maco Mlieč; Majster Bečko; Mamka Pôstková; Matka; Mátoha; Mišo; Moja matka; Na chlieb; Na hodoch; Našiel sa; Nové časy; Nový život; Obrázky; Ondrej; Pani urodzená; Pastierča; Pavúkovci; Pán syn; Po devätnásť; Počiatok konca; Pomník; Potrestaný úmysel; Predmluva; Prvé hodinky; Prvé prózy; Robotná knižka; Rodný dom; Sľuby; Smädni; Smrť Ďurka Langsfelda; Sokolská rodina; Spokojní; „Starý...“; Statky-zmätky; Stratená statočnosť; Strminou; Susedovo prasa; Štedrý večer; Tajní boháči; Ťažký boj; Tma; Úhrabky; Umrel Tomášik; Úžerník; V službe; Verše mladosti; Vypadly z hniezda; Z Čadce do mesta; Za rukopismi Záborského; Za tristo; Zaháľáč; Zajac; Zakrývame sa; Zrušil slovo; Zuza; Ženský zákon; Žliebky
- Urbánek, Ferko (1859–1934) Báj ružosadu; Beh žitia; Boh; Desatoro čitateľom; Diabol v raji manželskom; Hriešnica; Janko a Anička; Jánošíkov poklad; Kamenný chodníček; Kikimora; Kliatba; Kliatba a požehnanie; Krutohlavci; Mam a klam alebo Falošné diamanty; Moja vlasť; Ó, tie ženy!; O polnoci; Oštep a meč; Obrázky z hôr; Ozvena našich hôr; Pani richtárka; Pánik; Pokuta za hriech; Pytlíakova žena; Recitácie; Sólóvé výstupy pre dámy; Sólóvé výstupy pre pánov; Sňahulienka; Scény; Sen o šťastí; Sionitka; Slepý pastier; Slobodní manželia; Strašidlo; Stridža zpod Hája; Škriatok; Tatrín; Už sú všetci v jednom vreci; Za slobodu. Dramatický fragment; Zora; Ztratený život; Život
- Vadkerti-Gavorníková, Lýdia (1932–1999) Hra na pár-nepár; Kameň a džbán; Kolovrátok; Piesočná pieseň; Pohromnice; Totožnosť; Trvanie
- Vansová, Terézia (1857–1942) Chovanica; Danko a Janko; Hojže Bože!; Ján Vansa (Výber); Johankin zajac; Julinkin prvý bál; Kar; Kliatba; Magdalena; Matky; Mílka – mašamódkou; Milku dajú na edukáciu; Nové šatočky; Obete márnomyseľnosti; Ohlášky; Paľko Šuška; Pani Georgiadesová na cestách; Pani veľkomožná; Pút za šťastím; Prsteň; Recepty prastarej matere. Nová kuchárska kniha; Rozsobášení; Sestry; Sirota Podhradských; Stará pieseň; Šapšanko;

Terézia Medvecká, rodená Langeová; Z našej dediny
Zvon, Peter (1913–1942) Modlitba k umeniu; Tanec nad plačom

Slovak Newspapers and Magazines

Kankán: January to December 1997
Korzo: 2.5.1996 – 10.9.1996 and 2.9.1997 – 9.10.1997
Magazín večer: 21.2.1997
Práca: 1.6.1998-30.10.1999
Slovenská republika: 25.8.1997--20.9.1999
Sme Online: 26.7.2000--2.8.2000

Czech Corpus Texts

Technical Regulations

<http://ec.europa.eu/enterprise/tris/>
2005-0160.CZ, 2005-0666-CZ, 2005-0671-CZ, 2006-0161-CZ, 2007-0013-CZ, 2007-0135-CZ, 2007-0136-CZ, 2007-0137-CZ, 2007-0138-CZ, 2007-0357-CZ, 200500513.CZ, 200500525.CZ, 200500793.CZ, 200500815.CZ, 200501592.CZ, 200600816.CZ, 200601803.CZ, 200601851.CZ, 200602130.CS, 200602132.CS, 200602134.CS, 200700100.CZ, 200700261.CZ, 200700641.CZ, 200700643.CZ, 200700645.CZ, 200700647.CZ, 200701493.CZ

Czech Fiction

Borovský, Karel Havlíček (1821—1856) Král Lávro; Křest sv. Vladimíra; Tyrolské elegie;
Čapek, Karel (1890—1938) Adam stvořitel; Anglické listy; Básnické počátky; Bílá nemoc; Boží muka; Cesta na sever; Dášenska čili život štěněte; Devatero pohádek; Dodatky ke svazkům XIV-XIX; Hordubal; Hovory s T. G. Masarykem; Italské listy; Jak se co dělá; Jak vzniká divadelní hra; Kalendář; Kniha apokryfů; Korespondence 1; Korespondence 2; Krakatit; Krakonošova zahrada; Kritika slov; Lásky hra osudná; Loupežník; Marsyáš; Matka; Měl jsem psa a kočku; Menší prózy; O umění a kultuře 1; O umění a kultuře 2; O umění a kultuře 3; O věcech obecných čili Zoon politikon; Obrázky z Holandska; Obyčejný život; Od člověka k člověku 1; Od člověka k člověku 2; Od člověka k člověku 3; Povětroň; Povídky z druhé kapsy; Povídky z jedné kapsy; Překlady; První parta; R.U.R.; Skandální aféra Josefa Holouška; Továrna na absolutno; Trapné povídky; Univerzitní studie; Válka s mlouky; Věc Makropulos; Výlet do Španěl; Zahradníkův rok; Zářivé hlubiny a jiné prózy; Ze života hmyzu; Život a dílo skladatele Foltýna;
Čech, Svatopluk (1846–1908) Nový epochální výlet pana Broučka;
Celakovský, Ladislav (1799–1852) Ohlas písní českých; Ohlas písní ruských;
Dalimilova kronika;
Dyk, Viktor (1877–1931) Krysař;
Erben, Karel Jaromír (1811–1870) Jabloňová panna; Kytice;
Jirásek, Alois (1851–1930) Filozofská historie; Lucerna; Muzikantská Liduška; Písňe otroka; Po nás ať přijde potopa; Psohlavci; Staré pověsti české;
Karafiát, Jan (1846-1929) Broučci;
Klicpera, Václav Kliment (1792–1859) Divotvorný klobouk;
Klima, Ladislav (1878–1928) Utrpení knížete Sternenhocha;
Kolektiv (-) Máj Literární almanach Umělecké besedy; Památník českých žen;
Kořenský, Josef L. (1847–1938) Květena jarní;
Kulda, Beneš Metod (1820-1903) Čertův mlýn; O třech holubicích; Silný Jura;
Macha; Karel Hynek (1810–1836) Básně bez nadpisu a zlomky; Básně příležitostné; Básně

rozličné; Cikáni; Denník na cestě do Itálie; Denník z r. 1835; Dramatické zlomky; Jitro; Karlův Týn; Klášter sázavský; Korespondence Máchova; Krivoklad; Krkonošská pouť; Máchovy dopisy; Máj; Marinka; Mnich; Prvotiny; Valdek; Valdice; Večer na Bezdězu; Z Máchova denníku; Zlomky rozličné;

Mikšíček, Matěj (1815–1892) Ospalý Janek;

Mrštíkové, Alois a Vilém (1861–1925, 1863–1912) Maryša;

Němcová, Božena (1820–1862) Babička; Básně a jiné práce; Čert a káča; Čertův švagr; Chytrá horákyňe; Listy 1; Listy 2; Listy 3; Listy 4; Mahulena krásná panna; Národní báchorky 1; Národní báchorky 2; Národopis Česko; Národopis Slovensko; Neohrožený Mikeš; O Dlouhém, Širokém a Žárookém; O dvanácti měsíčkách; O hloupém Honzovi; O pánubohu; O Popelce; O princezně se zlatou hvězdou na čele; O ptáku ohniváku a mořské panně; O slunečniku měsíčníku a větrníku; O Smolíčkovi; O zlatém kolovrátku; O Zlatovlásce; Pohádka o kohoutkovi a o slepicce; Pohádka o perníkové chaloupce; Nemcova Pohorská vesnice; Povídky 1; Povídky 2; Princ Bajaja; Slovenské pohádky 1; Slovenské pohádky 2; Sůl nad zlato; Sedmero krkavců;

Neruda, Jan (1834–1891) Hřbitovní kvítí; Písně kosmické; Povídky malostranské; Zpěvy páteční;

Rais, Karel Václav (1859–1926) Zapadlí vlastenci;

Říha, Václav (1867–1937) Kovář Paška; Zvířátka a Petrovští;

Šalda, F. X. (1867–1937) Mácha snivec i buřič;

Stroupežnický, Ladislav (1850–1892) Naši furianti;

Světlá, Karolína (1830–1899) Černý Petříček; Kříž u potoka;

Těsnohlídek, Rudolf (1882–1928) Liška bystrouška;

Třebízský, Václav Beneš (1849–1884) Divotvorný ubrousek;

Tyl, Josef Kajetán (1808–1856) Strakonický dudák; Tvrdohlavá žena;

Winter, Zikmund (1846–1912) Mistr Kampanus;

Wolker, Jiří (1900–1924) Host do domu; Těžká hodina;

Vrchlický, Jaroslav (1853–1912) Noc na Karlštejně;

Zeyer, Julius (1841–1901) Jan Maria Plojhar; Tři legendy o krucifixu.

Translated Fiction into Czech

Aischylos -- Upoutaný Prometheus;

Hans Christian Andersen -- Císařovy nové šaty; Císařův slavík; Křesadlo; O hloupém Honzovi;

Sir Arthur Conan Doyle -- Mstitel; Návrat Sherlocka Holmese; Pes baskervillský; Poslední poklona Sherlocka Holmese; Příběhy Sherlocka Holmese; Údolí strachu; Vzpomínky na Sherlocka Holmese;

Viktor Hugo -- Básně;

Rudyard Kipling -- Kniha džunglí;

Friedrich Schiller -- Vilém Tell;

William Shakespeare -- Hamlet; Romeo a Julie

Czech Newspapers and Magazines

Bydlení: 21.10.1997 – 20.1.1998

Dáma: 11.10.1997 – 16.1.1998

Klenotník a hodinář: 15.10.1997 – 15.12.1997

Lidové noviny: no 287 – 294 1992 and 18.5.1996 – 27.1.1998

Magazín Lidových novin: 7.11.1997 – 26.1.1998

Metro: 25.11.1997 – 16.3.1999

Mladá fronta Dnes: 21.5.1996 – 3.10.1998

Náš chov: 15.10.1997 – 15.12.1997

Právo: 1.11.1997 – 27.1.1998

Strategie: 23.10.1997 – 18.12.1997

Strategie News: 12.11.1997 – 22.1.1998
Týden: 10.11.1997 – 26.1.1998

References

- Bąk, P. 1984. *Gramatyka języka polskiego*. 4th ed. Warszawa: Wiedza Powszechna.
- Bańko, M. 2002. *Wykłady z polskiej fleksji*. Warszawa: Wydawnictwo Naukowe PWN.
- & M. Krajewska. 1994. *Słownik wyrazów kłopotliwych*. Warszawa: Wydawnictwo Naukowe PWN.
- Bartek, H. 1944. *Správna výslovnosť slovenská*. Bratislava & Martin: Ján Horáček.
- Bartnicka, B., B. Hansen, W. Klemm, V. Lehmann & H. Siatkiewicz. 2004. *Grammatik der Polnischen*. München: Otto Sagner.
- Bartoš, J. 1972. *Grammaire de la langue slovaque*. Bratislava: Matica slovenská.
- Bartoš, F. 1901. *Nová rukověť správné češtiny*. Telč: Emil Šolc.
- Bauernöppel, J., H. Fritsch & B. Bielefeld. 1970. *Kurze tschechische Sprachlehre*. 2nd ed. Berlin: Vol und Wissen.
- Bělič, J., A. Jedlička, E. Jóna, E. Pauliny, J. Ružička & J. Štolc. 1980. *Slovenština*. 6th ed. Praha: Státní pedagogické nakladatelství.
- Benni, T., J. Łoś, K. Nitsch, J. Rozwadowski & H. Ułaszyn. 1923. *Gramatyka języka polskiego*. Kraków: Polska Akademia Umiejętności.
- Bethin, C. Y. 1992. *Polish syllables. The Role of Prosody in Phonology and Morphology*. Columbus: Slavica Publishers.
- . 1998. *Slavic prosody. Language change and phonological theory*. Cambridge: Cambridge University Press.
- Biedrzycki, L. 1974. *Abriß der polnischen Phonetik*. Warszawa: Wiedza Powszechna.
- Bielec, D. 1998. *Polish: An Essential Grammar*. London: Routledge.
- Brockhaus, W. 1995. *Final Devoicing in the Phonology of German*. Linguistische Arbeiten 336. Tübingen: Max Niemeyer Verlag.
- Bujalka, A., P. Baláž & A. Rýzková. 1998. *Slovenský jazyk I. Zvuková stránka jazyka. Náuka o slovnej zásobe*. Bratislava: Univerzita Komenského.
- Bujalka, A. & J. Dubníček. 1998. *Slovenský jazyk II. Morfológia*. Bratislava: Univerzita Komenského.
- Charette, M. 1990. Licence to govern. In: *Phonology* 7:2. 233-253.
- . 1991. *Conditions on Phonological Government*. Cambridge: Cambridge University Press.
- Cyran, E. 1994. *Vocalic Elements in Phonology. A Study in Munster Irish*. Lublin: Katolicki Uniwersytet Lubelski.
- . 2010. *Complexity Scales and Licensing in Phonology*. Studies in Generative Grammar 105. Berlin & New York: De Gruyter Mouton.
- & M. Nilsson. 1998. The Slavic [w > v] shift: a case for phonological strength. In: Cyran (ed.). *Structure and Interpretation. Studies in Phonology*. PASE Studies and Monographs. Volume 4. Lublin: Wydawnictwo Folium. 89-100.

- Damerau, N. 1967. *Polnische Grammatik*. Berlin: Walter de Gruyter & Co.
- Dejna, K. 1993. *Dialekty polskie*. 2nd ed. Wrocław: Ossolineum.
- Diklić, Z. (ed.). 1979. *Priručna gramatika hrvatskoga književnog jezika*. Zagreb: Školska knjiga.
- Długosz-Kurczabowa, K. & S. Dubisz. 2001. *Gramatyka historyczna języka polskiego*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Doroszewski, W. 1964. *O kulturę słowa*. 2nd ed. Warszawa: Państwowy Instytut Wydawniczy.
- . 1968. *O kulturę słowa. Poradnik językowy. Tom II*. Warszawa: Państwowy Instytut Wydawniczy.
- . 1979. *O kulturę słowa. Poradnik językowy. Tom III*. Warszawa: Państwowy Instytut Wydawniczy.
- Dubisz, S. (ed.). 1999. *Nauka o języku dla polonistów*. 3rd ed. Warszawa: Książka i Wiedza.
- Dukiewicz, L. 1995. Fonetyka. In: Wróbel (ed.). *Gramatyka współczesnego języka polskiego, Fonetyka i fonologia*. Kraków: Wydawnictwo Instytutu Języka Polskiego Polskiej Akademii Nauk. 7-103.
- EJP'91 = Urbańczyk, S. 1991. *Encyklopedia języka polskiego*. Wrocław: Ossolineum.
- Ertl, V. 1920. *Gebauerova krátká mluvnice česká*. 7th ed. Praha: Česká grafická unie.
- . 1926. *Gebauerova česká mluvnice pro školy střední a ústavy učitelské. I. Hláskosloví, Nauka o slově*. 9th corrected ed. Praha: Česká grafická unie.
- Falkenhahn, V. & W. Zielke. 1957. *Grammatik der polnischen Sprache*. Berlin: Volk und Wissen.
- Findra, J., G. Gotthardová, J. Jacko & E. Tvrdoň. 1986. *Slovenský jazyk a sloh*. 2nd ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- Gaertner, H. & A. Passendorfer. 1950. *Poradnik gramatyczny*. 2nd ed. Wrocław & Warszawa: Książnica - Atlas.
- Gebauer, J. 1904. *Příruční mluvnice jazyka českého pro učitele a studium soukromé*. 2nd corrected ed. Praha: Česká grafická společnost Unie.
- . 1905. *Mluvnice česká pro školy střední a ústavy učitelské*. 4th corrected ed. Praha: Česká grafická akc. společnost Unie.
- Gammelgaard, K. 1997. *Spoken Czech in literature. The case of Bondy, Hrabal, Placák and Topol*. Acta humaniora 18. Oslo: Scandinavian University Press.
- Goldsmith, J. 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell.
- Grappin, H. 1944. *Introduction phonétique à l'étude de la langue polonaise*. Paris: Librairie Droz.
- . 1963. *Grammaire de la langue polonaise*. 3rd ed, Paris: Institut d'études slaves.
- Gregor, A. 1932. *Mluvnice jazyka českého pro vyšší třídy středních škol*. Praha: Bursík a Kohout.

- Gruszczyński, W. & J. Bralczyk. 2002. *Słownik gramatyki języka polskiego*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.
- Gussmann, E. 2007. *The Phonology of Polish*. Oxford: Oxford University Press.
- & J. Kaye. 1993. Polish Notes from a Dubrovnik Café. In: *SOAS Working Papers in Linguistics and Phonetics* 3. London: SOAS, University of London. 427-462.
- Hála, B. 1962. *Uvedení do fonetiky češtiny na obecně fonetickém základě*. Praha: Nakladatelství Československé akademie věd.
- (ed.). 1967. *Výslovnost spisovné češtiny I. Zásady a pravidla. Výslovnost slov českých*. 2nd ed. Praha: Academia.
- Haller, J. 1934. O vokalizaci předložek v nové češtině. In: *Naše řeč* 18: 6-7. 178-187.
- Handbook of IPA. 1999. = *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Harris, J. 1990. Segmental complexity and phonological government. In: *Phonology* 7: 2. 255-300.
- . 1994. *English Sound Structure*. Oxford: Blackwell.
- & E. Gussmann. 1998. Final codas: why the west was wrong. In: Cyran (ed.). *Structure and Interpretation – Studies in Phonology. PASE Studies and Monographs* 4. Lublin: Folium. 139-162.
- Havlík, A. 1889. K otázce jerové v staré češtině. *Listy Filologické* 16: 1. 45-51.
- Havránek, B. & A. Jedlička. 1963. *Česká mluvnice*. 2nd corrected and expanded ed. Praha: Státní pedagogické nakladatelství.
- & ———. 1980. *Stručná mluvnice česká*. 15th ed. Praha: Státní pedagogické nakladatelství.
- & ———. 1981. *Česká mluvnice*. 4th revised ed. Praha: Státní pedagogické nakladatelství.
- Hlavsa, Z. et al. 1996. *Český jazyk pro střední odborné školy a studijní obory středních odborných učilišť všech typů*. Praha: SPN.
- Horák, G. 1999. *Slovo o slove*. Martin: Matica slovenská.
- Hůrková, J. 1995. *Česká výslovnostní norma*. Praha: Scientia.
- IJP'2012 = Vokalizace předložek (v třetí – ve třetí). In: *Internetová jazyková příručka. Ústav pro jazyk český*. <<http://prirucka.ujc.cas.cz/?id=770>> Last visited 13.08.2012.
- ISJP'00 = Mirosław Bańko (ed.). 2000. *Inny słownik języka polskiego I-II*. Warszawa: Wydawnictwo Naukowe PWN.
- Janda, L. A. & C. E. Townsend. 2002. *Czech*. SEELRC.
- Kačala, J. (ed.) 1976. *Kapitoly o slovenčine*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Kaleta, Z. 1995. *Gramatyka języka polskiego dla cudzoziemców*. Kraków: Uniwersytet Jagielloński.
- Kaye, J. 1990. Government in Phonology. The Case of Moroccan Arabic. In:

- Huybregts, Koster & van Riemsdijk (eds.). *The Linguistic Review*, Volume 6, 1986/1987, Number 1, Dordrecht: Foris. 131-159.
- . 1992. Do you believe in magic? The story of s+C sequences. In: *SOAS Working Papers in Linguistics and Phonetics* 2. 293-313.
- & J. Lowenstamm. 1981. Syllable Structure and Markedness Theory. In: Belletti, Brandi & Rizzi (eds.). *Theory of Markedness in Generative Grammar. Proceedings of the 1979 Glasgow Conference*. Pisa: Scuola Normale Superiore di Pisa. 287-315.
- , ——— & J.-R. Vergnaud. 1985. The internal structure of phonological elements: a theory of charm and government. In: Ewen & Anderson (eds.). *Phonology Yearbook* 2. 305-328.
- , ——— & ———. 1989. Konstituentenstruktur und Rektion in der Phonologie. In: M. Prinzhorn (ed.). *Phonologie. Linguistische Berichte, Sonderheft* 2. Westdeutscher Verlag. 193-231.
- Klebanowska, B., W. Kochański & A. Markowski. 1985. *O dobrej i złej polszczyźnie*. Warszawa: Wiedza Powszechna.
- Klemensiewicz-Bajerowa, I. 1952. O podwajaniu polskich przyimków. In: *Język Polski* 32:4, 156-163.
- . 1955. Polska oboczność *e* || *ø* w przyimkach i przedrostkach. In: *Studia z filologii polskiej i słowiańskiej* 1. Warszawa: Państwowe Wydawnictwo Naukowe. 162-202.
- Klemensiewiczówna, I. 1950. Czy wolno nam mówić *we wodzie ze sokiem*? In: *Język polski* 30:5, 193-203.
- Kočiš, F. 1990. *Slovenčina moja*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Kolaja, M. 1947. *Mluvnice a učebnice jazyka českého*. Praha: Státní nakladatelství.
- Kołodziejek, E. 1998. *Poprawna polszczyzna w praktyce*. Szczecin: Wydawnictwo Zachodniopomorskiej Szkoły Biznesu.
- Kott'1878 = F. Kott. 1878-1893. *Česko-německý slovník I-VII*. Praha: Josef Kolář.
- Kowalik, K. 1999. Morfonologia. In: R. Gregorczykova, R. Laskowski & H. Wróbel (eds.). *Gramatyka współczesnego języka polskiego. Morfologia*. Warszawa: Wydawnictwo Naukowe PWN, pp. 87-123.
- Kráľ, Á. 1988. *Pravidlá slovenskej výslovnosti*. 2nd ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- & J. Sabol. 1989. *Fonetika a fonológia*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Kraska-Szlenk, I. 2003. *The Phonology of Stress in Polish*. LINCOM Studies in Slavic Linguistics 23. München: LINCOM.
- Křčmová, M. 1984. *Fonetika a fonologie českého jazyka*, Praha: Státní pedagogické nakladatelství.
- KSSJ'97 = J. Kačala & M. Pisárčiková (eds.). 1997. *Krátky slovník slovenského jazyka*. 3rd ed. Bratislava: VEDA.
- Kucała, M. 1994. *Twoja mowa cię zdradza. Regionalizmy i dialektyzmy języka*

- polskiego*. Kraków: Towarzystwo Miłośników Języka Polskiego.
- Kučera, H. 1961. *The Phonology of Czech*. Slavistic Printings and Reprintings XXX. 's-Gravenhage: Mouton & Co.
- & G. K. Monroe. 1968. *A comparative quantitative phonology of Russian, Czech, and German*. New York: Elsevier.
- Kučera, K. 1982. K vokalizaci neslabičných předložek v češtině a slovenštině. In: *Acta Universitatis Carolinae, Philologica* 4-5, *Slavica Pragensia* XXV, 55-59.
- . 1984. K vokalizaci neslabičných předložek v současné češtině. In: *Náše řeč* 67:5, 225-237.
- Kuchař, J. & F. Váhala. (eds.). 1959. *Jazykový koutek Československého rozhlasu. Třetí výběr*. Praha: Československá akademie věd.
- Laskowski, R. 1975. *Studia nad morfonologią współczesnego języka polskiego*. Wrocław: Ossolineum.
- Leben, W. 1973. *Suprasegmental Phonology*. PhD dissertation, MIT. New York: Garland, published 1980.
- Letz, B. 1937. Vokalizácia predložiek. In: *Slovenská reč* 6, 191-197.
- . 1950. *Gramatika slovenského jazyka*. Bratislava: Štátne nakladateľstvo.
- Łoś, J. 1922. *Gramatyka polska. Część I. Głosownia historyczna*. Lwów, Warszawa, Kraków: Wydawnictwo Zakładu Narodowego imienia Ossolińskich.
- Mašín, J. 1928. *Česká mluvnice a cvičebnice pro I. a II. třídu středních škol. Díl I*. Praha: Profesorské nakladatelství a knihkupectví.
- Mazon, A. 1921. *Grammatire de la langue tchèque*. Paris: Librairie ancienne Honoré Champion.
- Meillet, A. & H. de Willman-Grabowska. 1921. *Grammaire de la langue polonaise*. Paris: Librairie Champion.
- Melichar, J. & V. Styblík. 1981. *Český jazyk. Přehled učiva základní školy*. 6th ed. Praha: Státní pedagogické nakladatelství.
- Mistrík, J. 1971. *Slovenčina pre každého*. 2nd ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- . 1983. *A grammar of contemporary Slovak*. Bratislava: Slovenské pedagogické nakladateľstvo.
- . 1984. *Moderná slovenčina*. Bratislava: Slovenské pedagogické nakladateľstvo.
- . 1994. *Gramatika slovenčiny*. Bratislava: Slovenské pedagogické nakladateľstvo.
- , O. Škvareninová & K. Hegerová. 1997. *Praktická príručka slovenčiny*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Mluvnice '86 = J. Petr (ed.). 1986. *Mluvnice češtiny 1. Fonetika, Fonologie, Morfologie a morfemika, Tvoření slov*. Praha: Academia.
- Mluvnice '95 = P. Karlík, M. Nekula & Z. Rusínová (eds.). 1995. *Příruční mluvnice češtiny*. Brno: Nakladatelství Lidové noviny.
- Mønnesland, S. 2002. *Bosnisk, kroatisk, serbisk grammatikk*. Oslo: Sypress Forlag.

- MSJP'74 = S. Skorupka, H. Auderska & Z. Łempicka (eds.). 1974. *Mały słownik języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Muczkowski, J. 1849. *Mała gramatyka języka polskiego*. Kraków: Drukarnia Uniwersytecka.
- Nagórko, A. 1998. *Zarys gramatyki polskiej (ze słowotwórstwem)*. 3rd ed. Warszawa: Wydawnictwo Naukowe PWN.
- Naše řeč*. 1917—. Praha: Ústav pro jazyk český Akademie věd České republiky.
- Naughton, J. 2005. *Czech. An essential grammar*. London: Routledge.
- Navrátil, L. 2001. *Neohybné slovné druhy a citoslovčia*. Nitra: Enigma.
- Nilsson, M. 1996. Sylabizacja niesylabicznych przyimków w językach zachodniosłowiańskich, zwłaszcza w czeskim. In: *Studia z leksykologii i gramatyki języków słowiańskich. VI polsko-szwedzka konferencja slawistyczna, Mogilany, 1-3 października 1995. Prace Instytutu Języka Polskiego 99*. Kraków: Instytut Języka Polskiego, Polska Akademia Nauk. 109-120.
- . 1998. Podmienky vokalizácie predložiek v češtine a slovenčine. In: *Varia VII. Zborník materiálov zo VII. kolokvia mladých jazykovedcov*. Bratislava: Slovenská jazykovedná spoločnosť pri Slovenskej akadémii vied. 211-220.
- Novotný, J. et al. 1992. *Mluvnice češtiny pro střední školy*. 2nd ed. Praha: Fortuna.
- NSPP'99 = A. Markowski (ed.). 1999. *Nowy słownik poprawnej polszczyzny*. Warszawa: Wydawnictwo Naukowe PWN.
- Obuchova 2010. = E. C. Обухова. *Чешский язык. Справочник по грамматике*. Москва: Живой язык.
- Oravec, J. 1968. *Slovenské predložky v praxi*. Bratislava: Slovenské pedagogické nakladateľstvo.
- , E. Bajžíková & J. Furdík. 1988. *Súčasný slovenský spisovný jazyk. Morfológia*. 2nd ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- & V. Laca. 1980. *Príručka slovenského pravopisu pre školy*. 6th ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- Ostaszewska, D. & J. Tambor. 2000. *Fonetyka i fonologia współczesnego języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Palková, Z. 1994. *Fonetika a fonologie češtiny s obecným úvodem do problematiky oboru*. Praha: Karolinum.
- Papierz, M. 1994. *Krótką gramatyka języka słowackiego*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.
- Pauliny, E. 1971. *Krátká gramatika slovenská*. 4th ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- (ed.). 1972. *Čeština*. Bratislava: Slovenské pedagogické nakladateľstvo.
- . 1981. *Slovenská gramatika*. Bratislava: Slovenské pedagogické nakladateľstvo.
- , J. Ružička & J. Štolc. 1955a. *Slovenská gramatika*. 3rd ed. Martin: Vydavateľstvo Osveta.

- , ——— & ———. 1955b. *Slovenská gramatika a sloh*. Bratislava: Slovenské pedagogické nakladateľstvo.
- , ——— & ———. 1968. *Slovenská gramatika*. 5th ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- Peciar, Š. 1947. O užívaní slabičnej a neslabičnej podoby predložiek v, s, z, k v spisovnej slovenčine. In: *Slovo a tvar* **1**. 43-49.
- Pettersson, T. 1979. *Tjeckisk fonetik och fonologi*. Göteborg: Institutionen för slaviska språk, Göteborgs universitet.
- Pisárčiková, M. et al. 1988. *Jazyková poradňa odpovedá*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Podracki, J. (ed.). 1993. *Polszczyzna płata nam figle*. Warszawa: Medium.
- . 2000. *Putapki językowe – w szkole i nie tylko...* Warszawa: JUKA.
- PSJČ'35 = *Příruční slovník jazyka českého I-VIII*. Praha: Státní nakladatelství/Státní pedagogické nakladatelství. 1935-1957.
- PSP'67 = Š. Peciar (ed.). 1967. *Pravidlá slovenského pravopisu*. 9th ed. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
- PSP'00 = M. Považaj (ed.). 2000. *Pravidlá slovenského pravopisu*. 3rd ed. Bratislava: VEDA.
- PSP'05 = I. Ripka, M. Imrichová & J. Skladaná. 2005. *Príručka slovenského pravopisu pre školy a prax*. Bratislava: Ottovo nakladateľstvo.
- PSWP'95 = H. Zgółkowa. (ed.). 1995–2005. *Praktyczny słownik współczesnej polszczyzny 1–50*. Poznań: Kurpisz.
- Romportl, M. 1985. *Základy fonetiky*. Praha: Státní pedagogické nakladatelství.
- Rowicka, G. 1999. *On Ghost Vowels*. The Hague: Holland Academic Graphics.
- Rubach, J. 1977. *Changes of consonants in English and Polish*. Wrocław: Ossolineum.
- . 1985. Lexical Phonology: lexical and postlexical derivations. In: *Phonology Yearbook* **2**. 157-172.
- . 1993. *The Lexical Phonology of Slovak*. The Phonology of the World's Languages. Oxford: Clarendon Press.
- Ružička, J., L. Dvonč, G. Horák, F. Miko, J. Mistrík, J. Oravec & M. Urbančok. 1966. *Morfológia slovenského jazyka*. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
- Ryba, B. 1934. Ve své domácnosti. In: *Naše řeč* **18**: 6-7. 173-178.
- Sabol, J. 1966. Používanie vokalizovaných predložiek **vo** a **ku** v spisovnej slovenčine. In: *Slovenská reč* **31**:5. 270-277.
- . 1989. *Syntetická fonologická teória*. Bratislava: Slovenská akadémia vied.
- Sawicka, I. 1974. *Struktura grup spółgłoskowych w językach słowiańskich*. Monografie Slawistyczne **27**. Wrocław: Ossolineum.
- . 1995. Fonologia. In: Wróbel (ed.). *Gramatyka współczesnego języka polskiego. Fonetyka i fonologia*. Kraków: Wydawnictwo Instytutu Języka Polskiego Polskiej Akademii Nauk. 105-195.
- Scheer, T. 1996. *Une théorie de l'interaction directe entre consonnes*. Thèse de

- doctorat. Université Paris 7.
- . 2000. *De la localité, de la morphologie et de la phonologie en phonologie*. Nice: Université de Nice Sophia-Antipolis.
- Sgall, P. & J. Hronek. 1992. *Čeština bez příkras*. Jinočany: H&H.
- Šimek, F. 1930. Faraonka. *Naše řeč* **14**: 9-10. 203-210.
- SJČ'41 = P. Váša & F. Trávníček (eds.). 1941. *Slovník jazyka českého*. 2nd extended ed. Praha: Fr. Borový.
- SJP 1900 = J. Karłowicz, A. Kryński & W. Niedźwiedzki (eds.). 1900-1927. *Słownik języka polskiego I–VIII*. Warszawa: Kasa im. Mianowskiego.
- SJP'58 = W. Doroszewski (ed.). 1958–1969. *Słownik języka polskiego I–XI*. Warszawa: Państwowe Wydawnictwo Naukowe.
- SJP'78 = M. Szymczak (ed.). 1978–1981. *Słownik języka polskiego I–III*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Šmilauer, V. 1972. *Nauka o českém jazyku*. Praha: Státní pedagogické nakladatelství.
- Sobol, E. 1996. *Podręczny słownik języka polskiego*, Warszawa, Wydawnictwo Naukowe PWN.
- Soerensen, A. 1900. *Polnische Grammatik in systematischer Darstellung*. Berlin: Langenscheidtsche Verlagsbuchhandlung.
- SOJP'86 = M. Szymczak (ed.). 1986. *Słownik ortograficzny języka polskiego wraz z zasadami pisowni i interpunkcji*. 10th ed. Warszawa: Państwowe Wydawnictwo Naukowe.
- SOPP'90 = S. Jodłowski & W. Taszycki. 1990. *Słownik ortograficzny i przepisy pisowni polskiej*. 13th ed. Wrocław: Zakład Narodowy imienia Ossolińskich.
- SPP'80 = W. Doroszewski (ed.). 1980. *Słownik poprawnej polszczyzny*. 2nd ed. Warszawa: Państwowe Wydawnictwo Naukowe.
- SSČ'94 = J. Filipec et al. (eds.). 1994. *Slovník spisovné češtiny pro školu a veřejnost*. 2nd rev. ed. Praha: Academia.
- SSJ'59 = Š. Peciar (ed.). 1959–1968. *Slovník slovenského jazyka I–VI*. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
- SSJČ'60 = B. Havránek (ed.). 1960–1971. *Slovník spisovného jazyka českého I–IV*. Praha: Nakladatelství Československé akademie věd.
- SSS'95 = M. Pisárčiková (ed.). 1995. *Synonymický slovník slovenčiny*. Bratislava: VEDA.
- SSSJ'06 = K. Buzássyová & A. Jarošová (eds.). 2006-2011. *Slovník súčasného slovenského jazyka A–L*. Bratislava: VEDA.
- Stanislav, J. 1953. *Slovenská výslovnosť. Príručka pre umelcov a iných verejných pracovníkov*. Martin: Osveta.
- . 1977. *Slovakische Grammatik*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Strutyński, J. 1995. *Gramatyka polska. Cz. 1. Wprowadzenie. Fonetyka. Fonologia*. Kraków: Księgarnia Akademicka.

- Styblík, V. et al. 1992. *Přehledná mluvnice češtiny pro základní školy*. Praha: Fortuna.
- Swan, O. 1980. *A Concise Grammar of Polish*. London: Orbis Books.
- . 2002a. *Polish Grammar in a Nutshell*. Pittsburgh: University of Pittsburg.
- . 2002b. *A Grammar of Contemporary Polish*. Bloomington: Slavica.
- SWJP'96 = B. Dunaj (ed.). 1996. *Słownik współczesnego języka polskiego*. Warszawa: Wilga.
- Szober, S. 1959. *Gramatyka języka polskiego*. 5th ed. Warszawa: Państwowe Wydawnictwo Naukowe.
- Szpyra, J. 1989. *The Phonology–Morphology Interface*. London: Routledge.
- . 1995. *Three tiers in Polish and English phonology*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Toporišič, J. 2000. *Slovenska slovnica*. 4th, expanded ed. Maribor: Obzorja.
- Townsend, C. E. 1990. *A description of spoken Prague Czech*. Columbus: Slavica Publishers.
- . & L. A. Janda. 1996. *Common and comparative Slavic: phonology and inflection with special emphasis on Russian, Polish, Czech, Serbo-Croatian, Bulgarian*. Columbus: Slavica Publishers.
- Trávníček, F. 1930. *Gebauerova příruční mluvnice jazyka českého pro učitele a studium soukromé*. 4th revised ed. Praha: Česká grafická unie.
- . 1949. *Mluvnice spisovné češtiny. Část I. Hláskosloví, tvoření slov, tvarosloví*. 2nd ed. Praha: Melantrich.
- . 1951. *Mluvnice spisovné češtiny. Část I. Hláskosloví, tvoření slov, tvarosloví*. 3rd corrected and extended ed. Praha: Slovanské nakladatelství.
- Uhlár, V. 1960. *Slovenská gramatika*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Urbańczyk, S. (ed.). 1966. *Polszczyzna piękna i poprawna*. 2nd ed. Wrocław: Ossolineum.
- USJP'03 = S. Dubisz (ed.). 2003. *Uniwersalny słownik języka polskiego 1-4*. Warszawa: Wydawnictwo Naukowe PWN.
- Vachek, J. 1968. *Dynamika fonologického systému současné spisovné češtiny*. Studie a práce lingvistické 8. Praha: Academia.
- Vážný, V. 1948. *Stručná mluvnice slovenská pro Čechy*. 2nd ed. Praha: Československá společnost.
- Vintr, J. 1994. *Tschechische grammatik*. Wien: Pädagogischer Verlag.
- Vojteková, M. 2011. *Predložky v spisovnej slovnecine a polštine*. Prešov: Prešovská univerzita v Prešove.
- Wierzchowska, B. 1980. *Fonetyka i fonologia języka polskiego*. Wrocław: Ossolineum.
- Wiśniewski, M. 2001. *Zarys fonetyki i fonologii współczesnego języka polskiego*. 4th ed. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.
- Wróbel, H. 2001. *Gramatyka języka polskiego*. Kraków: Od Nowa.
- WSPA'06 = Linde-Usiekniewicz, J. (ed.). 2006. *Wielki słownik polsko-angielski*.

Warszawa: Wydawnictwo Naukowe PWN.

Yoshida, S. 1990. A government-based analysis of the 'mora' in Japanese. In: *Phonology* 7: 2. 331-351.

———. 1991. *Some aspects of governing relations in Japanese phonology*. London: School of Oriental and African Studies, University of London.

Zauner, A. 1958. *Praktická příručka slovenského pravopisu*. 2nd ed. Martin: Osveta.

Zenkl, P. 1911. *Rádce správné češtiny*. Praha: Obchodnický spolek Merkur.

Zotova 2008 = С. В. Зотова. Как произносить предлоги: в или во, с или со?
In: *Русская речь* 2008: 5. 56-64.

Sammanfattning på svenska

Doktorsavhandling i slaviska språk vid Göteborgs universitet, 2014.

Titel: Vowel–Zero Alternations in West Slavic Prepositions. A Corpus Based Investigation of Polish, Slovak and Czech.

Författare: Morgan Nilsson.

Språk: Engelska, med sammanfattning på svenska.

Institution: Institutionen för språk och litteraturer, Göteborgs universitet, Box 200, 405 30 Göteborg.

Nyckelord: Polska, slovakiska, tjeckiska, vokalisering, prepositioner, inskottsvokal, korpus, fonologi.

ISSN 2000-1134

ISBN 978-91-981198-3-1

Syftet med denna avhandling är att undersöka användningen av inskottsvokal i prepositioner i de tre största västslaviska språken, dvs. polska, slovakiska och tjeckiska, och att med utgångspunkt i Government Phonology formulera de övergripande fonologiska principer som styr huvudparten av sådana vokalinskott.

Många prepositioner i de slaviska språken har två konkurrerande former, som t.ex. tjeckiskans *v* / *ve* *i*. Det är sedan länge ett känt problem att de villkor som styr bruket av dessa båda former inte lätt låter sig formuleras, en svårighet som har lett till komplikationer inom både preskriptiv språkvård avsedd för infödda språkanvändare och inom undervisning i dessa språk för utlänningar.

Den föreliggande undersökningen utgår från omfattande korpusdata som i första hand studeras med kvantitativa metoder, men även i viss utsträckning kvalitativt. Data och tidigare forskningsresultat hämtas också från litteraturen i ämnet. Kontexterna för förekomsten av prepositionernas vokaliserade form definieras så exakt som möjligt, och de fonologiska principer som ligger till grund för de fonologiskt betingade inskottsvokalerna diskuteras och formuleras.

En viss del av vokaliseringarna har visat sig bero på lexikaliseringar, och de är därmed inte fonologiskt betingade.

Resultaten i denna avhandling gör det möjligt att förse både språkstudier och infödda med ny, mera precis och detaljerad information om det faktiska bruket av prepositionernas vokaliserade form. Avhandlingen innehåller även vissa intressanta språkspecifika observationer rörande bruket av inskottsvokaler ur ett komparativt och typologiskt forskningsperspektiv.

DEPARTMENT OF LANGUAGES AND LITERATURES
GOTHENBURG SLAVIC STUDIES

1. Lishaugen, Roar. *Speaking with a forked tongue. Double reading strategies in Romány tři mágu by Jiří Karásek ze Lvovic*. 2008. 251 pp.
2. Karlsohn, Irina, Morgan Nilsson & Nadezjda Zorikhina Nilsson (eds.). *Литература как миропонимание. Literature as a world view. Festschrift in honour of Magnus Ljunggren*. 2009. 390 pp.
3. Nilsson, Morgan & Nadezjda Zorikhina Nilsson (eds.). *Семантический спектр славянского вида. IV конференция Комиссии по аспектологии Международного комитета славистов, Гётеборгский университет, 10 июня-14 июня 2013 г. Тезисы = The semantic scope of Slavic aspect. Fourth conference of the International Commission on Aspectology of the International Committee of Slavists, University of Gothenburg, June 10-14, 2013. Abstracts*. 2013. 173 pp.