Auditory cues determine allomorphy:

Vocalized and non-vocalized prepositions in Czech

Kateřina Chládková

MA thesis General Linguistics

University of Amsterdam

July 15, 2009

Supervisor: prof. dr. Paul Boersma

Second advisor: dr. Wolfgang Kehrein

Table of contents

Abstract

This study explains vocalizations of non-syllabic prepositions in Czech. It overcomes the limitations of previous research – which only explained vocalizations in a limited number of environments (Kučera 1984), or presented analyses with an otherwise unattested syllabic pattern of ‘pre-syllables’ (Kučera 1961), or had to assume a derivational analysis to account for *some* of the prepositional forms (Rubach 2000) – by showing that *all phenomena* concerning prepositional vocalizations can be explained using the *simplest principles* widely attested in the language. The analysis presented in this study thus accounts for vocalization patterns in all possible environments, and it can do without any special syllabification strategies such as ‘pre-syllables’. Most importantly, the long-held belief that prepositional vocalizations occur to lessen speakers’ articulatory effort (Hruška 1984, Havránek & Jedlička 1981) is replaced by a more plausible explanation. The explanation proposed in this study is listener-oriented, *i.e.* it is claimed that when a vocalized preposition occurs it is for the listener to be able to recover the preposition. Interestingly, it is found that not only onset properties but also language-specific preferences for particular prosodic structures determine the vocalizations. Crucially, unlike the previously proposed explanations, this study also fully clarifies the causes of the widely attested between- and within-speaker variation in prepositional vocalizations in certain contexts. It is also shown that vocalizations in other than non-syllabic prepositions do not require a separate mechanism but can well be explained by the mechanisms presented in this paper. The grammar proposed in this study is formalized by a model that operates on five levels of representation between the morpheme and the articulatory form. The model is created within the framework of Bidirectional Phonetics and Phonology (Boersma 2007). All the claims that the present analysis makes are supported by results of learning simulations.

1 Introduction

This paper focuses on non-syllabic prepositions, a phenomenon found in many Slavic languages (e.g. Russian, Slovenian, Polish, Slovak, Czech). Non-syllabic prepositions can have the form of either a single consonant (e.g. *k domu* 'to a house') or can be *vocalized*, *i.e.* the consonant is followed by a 'fill' vowel (e.g. *ke kolu* 'to a bike'), and thus become syllabic. The present study is about the vocalized and non-vocalized forms of non-syllabic prepositions in Czech.

It has been widely acknowledged (Dickins 1998, Čermák 1996, Kučera 1984, Havránek & Jedlička 1981)[1] that there is a rather high variability in Czech with respect to when the vocalizations of non-syllabic prepositions occur and when they do not occur. Only few studies address this question directly (Dickins 1998, Kučera 1984), but no definitive generalization has been made available yet. It is crucial to give a more accurate and systematic explanation of the distribution of vocalized and non-vocalized non-syllabic prepositions, which represent almost 60 % of occurrences of all the Czech prepositions (Čermák 1996). This study aims at modeling the Czech speakers' grammar in order to fully account for the observed distribution of non-syllabic prepositions.[2] I will first present a set of generalizations, based on the phonological and morpho-phonological properties[3] of the prepositions and their post-contexts (*i.e.* words that they pre-modify), and then create a model of Czech speakers' grammar, which will be able to predict what we observe and, at the same time, will account for the variation that is indisputably present in the use of the vocalized and non-vocalized forms of the prepositions. In simulations with virtual learners, I will show that the proposed grammar is learnable.

---

[1] All translations of quotes from languages other than English are mine.

[2] Since most literature on the topic considers Bohemian Czech (the variety of Czech spoken in Bohemia, the western part of the Czech Republic) I will, too, consider and model the Bohemian variety in the present study. I will provide notes on the Moravian pronunciation (the variety spoken in the eastern part of the country) where marked and consistent differences are observed.

[3] I ignore all other factors such as idiomatization, speaking style, degree of clarity, or phonetic structure of *preceding* words (*i.e.* the nature of their final consonant); see Dickins (1998) for a discussion of these.

In Section 2, I give a description of the distribution of the four non-syllabic preposition pairs *k* – *ke*[4] ('to'), *v* – *ve* ('in'), *z* – *ze* ('from'), *s* – *se* ('with'). All observations reported in this study are based on the analysis of the SYN2005 subcorpus of the *Czech National Corpus*. A brief note on vocalization effects found in other, always-syllabic, prepositions such as *od* – *ode* ('from'), *bez* – *beze* ('without'), *před* – *přede* ('before') is made as well. Since vocalizations in the always-syllabic prepositions are extremely rare and seem to appear mostly in idiomatized expressions (which are ignored in this study), in the present analyses, attention is paid primarily to the non-syllabic prepositions and their vocalized forms. Crucially, the final model of the speakers' grammar that I create on the basis of the vocalizations observed in non-syllabic prepositions will account for the vocalization of the always-syllabic prepositions as well.

Section 3 gives an overview of studies that proposed generalizations and explanations for the prepositional vocalizations, with more or less explanatory power and/or accuracy.

Section 4 presents in detail the steps of formalizing the Czech grammar of prepositional vocalization and Section 5 describes a test of the grammar done by learning simulations and the results thereof.

The main findings of this study and suggestions for future work are summarized in Section 6.

[4] In the present study I ignore the other, rarely occurring, vocalized form of this preposition – *ku*. In contemporary Czech, it seems to be used only in a few idiomatized expressions (such as *ku prospěchu* 'to the benefit'); it used to occur before bilabial consonants (such as in *ku Praze* 'to Prague') but it nowadays sounds obsolete (as Kučera (1984) also observes).

2 Description of the data and generalizations

Prepositions in Czech pre-modify a noun phrase; the word that they immediately precede may thus be a noun, an adjective, a numeral, a pronoun, or an adverb. It seems that the nature of the onset cluster that results from linking the non-syllabic preposition to the word is a primary determiner of whether the preposition will take the vocalized form or not, while the prosodic structure of the whole sequence formed by the preposition and the immediately following word may as well play a role in assigning the form of the preposition.

2.1 Simple onsets: the clear-cut cases

The non-vocalized forms of the prepositions occur whenever the following word starts in a vowel (in its spelled form); in Czech, such words are realized with a glottal stop before the vowel (Hála 1962:280, 379).[5]

Next, if the preposition is followed by a simple onset, and the onset consonant differs at least in place or manner from the preposition, then the preposition takes its non-vocalized form. If the simple word onset is followed by a syllabic *r* or *l*, /r̩/, /l̩/, the prepositions take their non-vocalized form, too. Examples of prepositions followed by vowel-initial and single-consonant-initial words are given in (1a) and (1b), respectively; the spelled forms are listed[6]

[5] Importantly, note that this statement seems to hold only for the varieties of Czech spoken in Bohemia. In Moravia, if a vowel-initial word is preceded by a non-vocalized non-syllabic preposition, the glottal stop is usually not pronounced and the consonant of the preposition, if underlyingly voiceless, is realized as voiced. That is, the respective phonetic forms of the first four examples in (1) uttered by a speaker of the Moravian dialect will be /goknu/, /vokɲɛ/, /zokna/, /zoknɛm/. Hála (1962) judges this to be an incorrect pronunciation. However, this with no doubt is the norm in the Moravian variety of Czech. Palková (1994:326) notes this as well.

[6] Note that the Czech orthography is transparent (except for voicing). With respect to the presence or absence of vocalization it reflects the pronunciation. Therefore, whenever a vocalized form of a preposition is used, this will also be shown by the orthography, *i.e.* /kɛ/ and /k/ correspond to *ke* and *k*, respectively.

(first column) together with their meanings (second column) and their phonetic realizations (third column)[7].

(1a) VOWEL-INITIAL WORDS

k oknu	‘to the window’	/kʔoknu/
v okně	‘in the window’	/fʔokɲɛ/
z okna	‘from the window’	/sʔokna/
s oknem	‘with the window’	/sʔoknɛm/

(1b) SIMPLE-ONSET WORDS[8]

k domu	‘to a house’	/gdomu/
k choti	‘to a spouse’	/kxocɪ/
k Brnu	‘to Brno’	/gbr̩nu/
k místu	‘to a place’	/kmi:stu/
k vodě	‘to the water’	/kvoɟɛ/
v parku	‘in a park’	/fparku/
v zemi	‘in the ground’	/vzɛmɪ/
v mlze	‘in the mist’	/vml̩zɛ/
v místě	‘in a place’	/vmi:scɛ/

[7] The phonetic symbols used in this study are IPA.

[8] Note in the phonetic representations that the non-vocalic preposition assimilates in voicing to a following obstruent, but not to a following sonorant. This is because generally in Czech, obstruents but not sonorants trigger regressive assimilation. The preposition *s* is a special case as it can be both assimilated in voicing to the following sonorant or preserve its voicelessness. Another special case is the phoneme /v/, which does undergo regressive assimilation but is not able to trigger it itself, *i.e.* /fparku/ ‘in a park’ but /kvoɟɛ/ ‘to the water’; Palková (1994:328-331).

z domu	'from a house'	/zdomu/
z filmu	'from a movie'	/sfɪlmu/
z místa	'from a place'	/zmi:sta/
s domem	'with a house'	/zdomɛm/
s filmem	'with a movie'	/sfɪlmɛm/
s místem	'with a place'	/smi:stɛm/ or /zmi:stɛm/

If the preposition and the following onset consonant have identical place and manner speficications (irrespective of their underlying specifications for voicing[9]), the preposition always takes its vocalized form, for examples see (2a). I stress here the fact that it is the *place together with the manner* of the onset consonant that must be identical to the place and manner of articulation of the preposition to cause the preposition to vocalize. It has been abundantly but incorrectly noted by Dickins (1998) that the vocalized forms of prepositions occur when the following consonant is homorganic to the preposition (p. 210, 213, 214, 215, 228). The search of the SYN2005 corpus shows that there is, for instance, not a single example of a word beginning in a simple onset formed by /x/ that is preceded by the vocalized preposition *ke* (*i.e.* words such as *chata* /xata/ 'a summer house' or *chalupa* /xalupa/ 'a cottage' are always preceded by the non-vocalized preposition *k*). Similarly, any word beginning in a single alveolar plosive would always be pre-modified by the non-vocalized forms of the prepositions *s* and *z*; see the examples in (1b).

The corpus analysis shows, in line with what previous studies claimed (Kučera 1984), that the always-syllabic prepositions behave differently from the non-syllabic prepositions when the (final) consonant of the preposition is identical (after voice assimilation) to the onset of the following word. The identity between the two segments (*i.e.* between the final consonant of the preposition and the following onset) hardly causes vocalization of the always-syllabic prepositions, see (2b). In this paper I focus on non-syllabic prepositions but in

[9] After regressive voice assimilation in an obstruent cluster, the obstruents surface all as either voiced or voiceless.

the final section of this study I will show that eventually, in both the non-syllabic and always-syllabic prepositions, vocalizations are handled by the same principles, contrary to Kučera's (1984) claim that vocalizations in the two groups of prepositions are affected by different factors and should be dealt with separately.

(2a) IDENTICAL ONSETS: NON-SYLLABIC PREPOSITIONS

ke kolu	'to a bike'	/kɛkolu/
ke gólu	'to a goal'	/kɛgoːlu/
ve fyzice	'in physics'	/vɛfɪzɪt͡sɛ/
ve vesnici	'in a village'	/vɛvɛsɲɪt͡sɪ/
ze země	'from the ground'	/zɛzɛmɲɛ/
ze sýra	'from cheese'	/zɛsiːra/
se zemí	'with the ground'	/sɛzɛmiː/
se sýrem	'with cheese'	/sɛsiːrɛm/

(2b) IDENTICAL ONSETS: ALWAYS-SYLLABIC PREPOSITIONS

od domu	'from house'	/oddomu/[10]
bez sýra	'without cheese'	/bɛssiːra/

2.2 Complex onsets: high variability

In (3) I give a list of word onsets ordered from top – onsets that always occur with the vocalized preposition (*i.e.* identical onsets), to bottom – onsets that always occur with non-

[10] The Czech underlying phoneme inventory does not contain phonologically long consonants. Hála (1962:216) notes that on the surface, adjacent identical consonants are more or less tied together and can be phonetically realized as a slightly prolonged consonant; they merge into a phonetically single short consonant only in less careful speech.

vocalized preposition. The list is based on the tendencies observed in the SYN2005 corpus. The exact percentages are not given here because the corpus is not a spoken but a written corpus, which can partly affect the proportion of vocalizations, and neither can it provide data on between- and within-speaker variability. Moreover, not all word onsets that are possible in Czech are present in the corpus with each of the four prepositions. However, the tendencies are expected to reflect the actual situation that is attested in the spoken language, at least with respect to the onset-type dependent amount of prepositional vocalizations.

The list in (3) is divided into 3 parts, (3a) through (3c), corresponding to the four non-syllabic prepositions (*s* and *z* are listed together because they only differ in voicing, a feature in which they most often assimilate with the first consonant of the onset). Because the prepositions differ from each other in place and manner of articulation, it is not surprising that there are slight differences in which onsets are given in (3a), (3b), and (3c), and how they are ordered. Eventually, in Section 4.7 it will be seen that the generalizations drawn from (3a), (3b), and (3c) are similar irrespective of the particular features employed in determining the degree of the preposition-onset similarity.[11] The list in (3) is an overview of the general tendencies in the corpus, on the basis of which I will propose generalizations, which apply to the vocalizations of all the four non-syllabic prepositions.

(3)	PREPOSITION	FOLLOWING-WORD ONSET	VOCALIZED
(3a)	***k***	/k/, /g/, /kC/, /gC/	always
		/skr̥/, /skv/, /ʃkv/	
		/sk/, /ʃk/, /sx/	often

[11] Note that before the personal pronoun *já* ‘I’ in oblique cases (which all start in the onset /mɲ/ or /mn/) and traditionally also before the pronoun *všechen* /fʃɛxɛn/ ‘all’ the vocalized form of the prepositions, both the always-syllabic and non-syllabic, is always used; in this paper I ignore these cases as they have been considered to be lexicalized, *i.e.* the presence of the vocalized form does not seem to be due to the properties of the onset (Dickins 1998). Compare *ke mně* /kɛmɲɛ/ ‘to me’ and *k městu* /kmɲɛstu/ ‘to town’, *k měsíci* /kmɲɛsiːt͡sɪ/ ‘to moon’; *ode mě* /odɛmɲɛ/ ‘from me’ and *od města* /odmɲɛsta/ ‘from town’, *od měsíce* /odmɲɛsiːt͡sɛ/ ‘from moon’.

		/rt/, /rv/, /lɦ/, /lʒ/	
		/t͡sc/, /fst/, /fsp/	
		/st/, /sp/, /sf/	
		/xr/, /xr/, /xl/, /xv/, /ɦv/, /ɦr̥/	sometimes
		/vzn/, /dr̥/, /br̥/, /tr̥/, /dn/, /tm/, /ps/	
		/ml/, /pl/, /tr/, /pr/, /mr/	rarely
		/t/, /p/, /f/, /l/, /s/… etc.	never
(3b)	***v***	/f/, /v/, /fC/, /vC/	always
		/rv/, /lv/, /rt/, /lʒ/, /lɦ/	often
		/skv/, /ʃkv/, /skr̥/, /sf/, /zdv/	
		/sv/, /zv/, /xv/, /ɦv/	
		/sb/, /sp/, /sk/, /ʃk/, /zd/, /zb/	sometimes
		/dv/, /tv/, /ɦr̥/, /tr̥/, /dn/, /tm/, /ps/	
		/ml/, /pl/, /tr/, /pr/, /mr/	rarely
		/t/, /p/, /k/, /l/, /ʃ/… etc.	never
(3c)	***z, s***	/s/, /z/, /sC/, /zC/	always
		/rt/, /rv/, /lʒ/, /lɦ/	often
		/vz/, /fs/, /t͡sc/, /fʃ/	
		/ʒl/, /ʒv/, /ʒr/, /ʃl/, /ʃp/, /ʃk/, /ʃv/, /ʃr/	
		/tr̥/, /mzd/, /dn/, /tm/, /ps/	sometimes
		/ml/, /pl/, /tr/, /pr/, /mr/	rarely
		/t/, /p/, /k/, /l/, /f/… etc.	never

The list in (3) confirms that in words with simple onsets, as has been noted in the Section 2.1, absolute identity results in exclusive vocalization, and non-identity in exclusive non-vocalization. With complex onsets we encounter high variability: the greater the similarity of the segments in the resulting consonant cluster, which is created when the non-vocalized preposition is linked to the following onset, the more likely it is that the vocalization will occur. The vocalization is variable in all cases other than the case of identity between the onset-initial segment and the preposition, while the degree of optionality in assigning the vocalized or the non-vocalized form of a preposition varies greatly between and within speakers (Dickins 1998, Kučera 1984).

For instance, the onset /sk/ will always require prepositional vocalization when preceded by the preposition *s* or *z*; this is because of the identity between its first segment and the preposition. Next, when such an onset is preceded by the preposition *k*, vocalization will be required very often (almost always); this is because the second segment, *i.e.* /k/, is identical to the preposition; the resulting cluster formed by the preposition and the onset is thus /ksk/. In such a resulting cluster, there are 3 consonants, but only two of them are different. Last, with the preposition *v* the vocalization will sometimes occur, but it will occur less often than with the preposition *k*. This is because the resulting cluster is /fsk/; within this cluster no segment is identical in both its place and manner specification to any other segment.

The fact that more similar clusters require vocalized preposition more often than less similar clusters naturally implies that more complex onsets will tend to incorporate more similarity than less complex onsets. For instance, a word onset such as /skr̥/ will always contain more similarity than an onset such as /sk/ irrespective of whether it is preceded by the preposition *k* or *v*. In other words, both /kskr̥/ and /fskr̥/ contain more similarity than /ksk/ and /fsk/, respectively. In /kskr̥/ there are two identical consonants, and two consonants that are both alveolar fricatives, while in /ksk/ there are only two identical consonants. Similarly, in /fskr̥/ there are three voiceless fricatives, two of which have identical place features, while in /fsk/ only two fricatives have an identical place feature. It therefore seems that out of the definition of similarity, the complexity of the resulting onset emerges as a related factor that also contributes to the prepositional vocalization. Another example of the interaction between onset similarity and complexity is found in onsets that begin in /ʒ/ or /ʃ/ and are preceded by the prepositions *s* or *z*. If such an onset is simple, the vocalization of *s* and *z* is not at all obligatory[12] and it takes place considerably less often than if the onset contains a second C. This is because adding an extra alveolar segment in /ʒl/, /ʒr/, or /ʃl/ (as opposed to the simple /ʒ/ or /ʃ/) or an extra voiceless labial segment in /ʃp/ yields greater similarity within the resulting onsets such as /zʒl/, /zʒr/, /sʃl/ or /sʃp/. See section 4.7 for a more detailed analysis of these two related factors.

Interestingly, from the list in (3) it is seen that there is a type of onsets, which – irrespective of their similarity – very often bring about obligatory vocalizations. These onsets have the form sonorant – obstruent (SO), e.g. /rt/. The explanation that I propose is that the pre-modified word wants to preserve its syllabic pattern. It has been put forth that the non-syllabic non-vocalized prepositions become a part of the following syllable (Petr *et al*. 1986, Rubach 2000). If a word such as *rtuti* /.rtu.cɪ./ 'mercury-DAT' is preceded by a non-vocalized preposition (e.g. *k rtuti* 'to mercury'), there are two possibilities with respect to the syllabic status of the /r/. I will now describe the two possibilities and show that both of them are somehow problematic.

Either, the sonorant /r/, which is not syllabic in the word *rtuti* and is the first C of the onset /rt/, can become a second C of a OSO onset /.krtú.cɪ./. Or, the /r/ can become syllabic, thus forming a nucleus of a CV syllable, of which the preposition *k* is an onset /.kŕ̩.tu.cɪ./. The problem with the first scenario is that a language such as Czech, which has a syllabic /r/, does not allow OSO clusters where the S is non-syllabic[13] (such as in /.kr̩.tɛk./ 'a mole'). However, the /r/ in preposition-word sequences such as *k rtuti* is never realized as syllabic. If in the language a word such as *rtuti* is preceded by a non-vocalized preposition, which happens in the minority of cases, the /r/ stays non-syllabic. There thus seems to be a tendency in the language to preserve the consonantal status of this word-initial /r/. The easiest way to do this, then, is to vocalize the preposition, *i.e.* /.kɛ́.rtu.cɪ./.

[12] Kučera reports that exactly in 50% of these cases the preposition is vocalized.

[13] If the sonorant /l/ or /r/ (and in a few cases also /m/ and /n/) is between at least two consonants within a word, or is word-final and preceded by at least one consonant, then it is syllabic (Palková 1994:270).

Besides the aforementioned resulting cluster similarity (plus its related complexity), and the sonority contour, there is another factor that affects the form that the preposition will take: it is the prosodic structure of the whole resulting preposition-word sequence; how prosodic structure influences the vocalizations cannot be read off from the list in (3) but will be illustrated in (4).

Stress in Czech falls on the first syllable of a word. If a word is pre-modified by a monosyllabic primary preposition[14], the stress is shifted from the word's first syllable to the preposition (Palková 1994:338). The preposition and the immediately following word thus form a prosodic word with initial stress. Prosodic words in spoken Czech are mostly bisyllabic (42%) and trisyllabic (30%; Palková 1994:285). Monosyllabic words in Czech are well attested, but monosyllables that are formed by a light syllable[15] occur only rarely (Kager 1995).

Taking into account the properties of Czech described in the previous paragraph, I argue that the extent to which the vocalized form of a preposition is obligatory depends on the syllabic structure of the resulting prosodic word. From the examples in (4) it becomes clear that onset complexity and similarity are not the only factors affecting the vocalization. Monosyllabic words (with a complex onset) – most obligatorily if the monosyllable is light (compare e.g. the variation in *ke psům/k psům* and the obligatory vocalization in *ke psu*) – almost always require the vocalized preposition, as opposed to longer words that start with the same onset. In line with the well-attested high frequency of bisyllabic and trisyllabic prosodic words, and the extremely low frequency of light monosyllables in Czech, it seems that the preference *for* bisyllabic words and *against* light monosyllables is reflected in the forms the

[14] Czech has two classes of prepositions. Primary prepositions such as all those that are dealt with in the present paper, and secondary prepositions, such as *okolo* 'about', *místo* 'instead of'. The latter class, but not the former class, can carry full semantic meaning and can thus function as an adverb or a noun if standing alone. (Hála 1962:310). Furthermore, the stress in phrases with the secondary prepositions and with polysyllabic primary prepositions does not shift from the pre-modified word to the preposition, but usually both the preposition and the word are stressed.

[15] Light monosyllables are monomoraic, heavy syllables are at least bimoraic. A short vowel in a syllable nucleus represents one mora and a long vowel represents two moras. A syllable coda represents one mora.

prepositions take. If the Czech grammar needs to avoid having a light monosyllabic word, prepositions will be vocalized (e.g. *ke cti*).

(4)	*ke cti*	'to dignity'	/kɛt͡scɪ/
	k ctižádosti	'to ambition'	/kt͡scɪʒa:doscɪ/
	ke psu	'to dog'	/kɛpsu/
	ke psům (sometimes *k psům*)	'to dogs'	/kɛpsu:m/
	k psovi (sometimes *ke psovi*)	'to dog'	/kpsovɪ/
	se psy	'with dogs'	/sɛpsɪ/
	se psím (sometimes *s psím*)	'with dog (Adj., Sg.)'	/sɛpsi:m/
	s psími (sometimes *se psími*)	'with dog (Adj., Pl.)'	/spsi:mɪ/
	ze dne	'from a day'	/zɛdnɛ/
	z dneška	'from today'	/zdnɛʃka/

Interestingly, but after the above observation not very surprisingly, Trávníček (1949:50) notes that the non-vocalized forms occur more often before adjectives than before nouns. Hardly any adjectival form in Czech is monosyllabic, as opposed to monosyllabic nouns (in oblique cases), which are not uncommon in Czech. With the monosyllabic nouns such as *dne* 'a day-GEN', *dni* 'a day-DAT', *snu* 'a dream-DAT', *hry* 'a play-GEN', *mši* 'a mass-DAT', vocalized forms of the prepositions are almost exclusively used. It might therefore not be the category of the word, as Trávníček claims, but simply the length (*i.e.* the number and the weight of syllables) of the word that determines the vocalization.

2.3 Complex onsets: generalizations

The pattern of prepositional vocalization that we saw in the words with complex onsets can be summarized as follows.

The vocalization seems to be most frequent when the segments of the resulting preposition-onset cluster share a lot of similar features with each other; the more similar the Cs within the resulting cluster, the more frequent the vocalization. It was also shown that clusters composed of more segments naturally tend to contain more similarity than clusters with fewer segments.

It is worth noting that related polysyllabic and monosyllabic words with the same onset differ with respect to the frequency of prepositional vocalizations. In (4) it was seen that the vocalized preposition is always used if the complex-onset word is a light monosyllable, and almost always if it is a heavy monosyllable. I thus add another generalization, which concerns the structure of the prosodic word (which is formed by the preposition and the immediately following word), *i.e.* the requirement for it to be at least a heavy monosyllable or, most preferably, to be bisyllabic. I have shown that it is possible to find an explanation for the variability in the vocalizations if we do not restrict our analyses to the phonetic properties of the (resulting) cluster alone, but take into account the prosodic structure of the whole preposition-word sequence.

Based on the facts reported in the literature and on the observations made in the corpus (which correspond to the author's native-speaker intuition), I have proposed several major generalizations that could explain the distribution of vocalized and non-vocalized non-syllabic preposition in Czech. The four main predictors of the vocalizations thus seem to be the following:

(i) the degree of similarity in the resulting preposition-word onset cluster

(ii) the complexity of the resulting onset cluster,

(iii) the sonority contour of the resulting onset cluster (consonantal vs. vocalic identity of the segments),

(iv) the length of the whole resulting preposition-word sequence.

3 Previous research

The literature on Czech rather briefly notes that some prepositions occur in vocalized and in non-vocalized forms, but does not provide a detailed description of the contexts, in which we observe the vocalized or non-vocalized prepositions. Similarly, an explanation is missing for whether the two forms are two allomorphs underlyingly or whether the vowel is epenthesized or deleted during production or perception.

Several studies focused in more detail on describing and explaining vocalized and non-vocalized verbal prefixes, some of which are considered to have developed from prepositions; nevertheless these studies acknowledge that although the prefixes and prepositions might have had the same origin, their status and their behavior in contemporary Czech is very different (Ziková 2008).

3.1 Distribution of the prepositions

A few studies attempted to present an analysis of vocalizations in Czech prepositions (Dickins 1998, Kučera 1984). By means of a questionnaire to three informants who were (supposedly native) speakers that had “a special interest in Czech” (p. 202), Dickins (1998) aimed to provide a systematic account of when a preposition is vocalized and when it is non-vocalized. However, generalizations are difficult to find in his study, which eventually concluded, unsurprisingly and similarly to the statements in various grammars of Czech, that the only environments in which vocalized prepositions are obligatory are the personal pronoun *já* in its oblique cases (which all start in /mɲ/ or /mn/) and word onsets homorganic with the (final consonants of the) preposition – a conclusion that does not even seem to be correct: as I noted above, it is the identity of place *and* manner of articulation that yields obligatory vocalization (as also Kučera (1984) observes). Similar to the present study, at least one earlier study (Kučera 1984) focused on non-syllabic prepositions and reported percentages of the vocalized and non-vocalized prepositions drawn from raw language data. However, because this earlier study examined only prepositions followed by words with simple onsets, it does not offer a full account of the vocalizations. It was seen in Sections 2.1 and 2.2 that the vocalizations concerning simple onsets are rather straightforward, and that the greatest variation and complexity lies in the words with complex onsets.

The present study analyzes both words with simple onsets as well as words with complex onsets, and aims not only to answer the question under which circumstances the vocalizations occur but importantly, it aims to find out what causes the variation that previous studies reported on.

3.2 Proposed explanations

In the literature it has been almost exclusively assumed that the syllabic forms of the four prepositional meanings are used whenever speakers have the need to achieve articulatory ease. “Some Czech non-syllabic prepositions may be formed by one consonant sound only (e.g., v, k, s, z). These, however, may be changed to monosyllabic prepositions (ve, ke, ku, se, ze) before some consonant groups to make the pronunciation easier” (Hruška 1976:141). Similarly, according to Havránek & Jedlička (1981) the vocalized form occurs in order to make the articulation easier when the following word begins in a consonant identical or similar to the preposition, or if the onset is CCC; before CC onsets both vocalized and non-vocalized prepositions occur. Kučera (1984) sums up that the only reason for prepositional vocalizations reported in the literature is articulatory ease. The grammars of Czech and all the previous studies seem to assume that the non-vocalized form is the underlying form, which during the process of production eventually surfaces with an /ɛ/ to facilitate the articulation.

I argue that it is highly unlikely that articulatory constraints would favor an addition of an extra segment to minimize the articulatory effort. A full vowel such as /ɛ/ requires an extra articulatory effort (e.g. tongue and jaw movements, setting vocal folds in vibration) that does not have to be present if the preposition is non-vocalized. There does not seem to be an articulatory constraint that would prevent the Czech speakers from pronouncing clusters such as /ksk/, because they are well used to pronouncing clusters such as /ks/ (e.g. *ksicht* ‘face, Coll.’ or *komplex* ‘a complex’) as well as clusters such as /sk/ (e.g. *skok* ‘a jump’ or *lesk* ‘a shine’), or even clusters like /pstr/ (e.g. *pstruh* ‘trout’). Moreover, and perhaps even more convincingly, pronouncing the phrase *k kolu* ‘to a bike’ as [kːolu] undoubtedly requires less articulatory effort than [kkolu], which in turn requires less articulatory effort than [kɛkolu]; this is because in [kːolu] the speaker would not have to realize an extra vowel [ɛ] and could also realize only one plosive (prolonged) instead of two. It is clear that if articulatory constraints were to predict the prepositional vocalizations, they would favor the form [k:olu], which is not attested.

To my knowledge, all studies except one have adopted the – as was just seen – unlikely articulatory explanation. An exception is Rubach (2000) who in his Derivational Optimality Theoretic analysis of several phenomena found in Slavic languages notes that geminates (such as those that would occur if 'to a bike' were realized as /kkolu/) are prohibited in Czech by the Obligatory Contour Principle (OCP), which bans structures with identical adjacent segments (e.g. McCarthy 1986). This structural OCP is dominated by a faithfulness constraint against deleting segments, and thus /ɛ/ is inserted.

In the present study, the notion of OCP will be dealt with as well but instead of being structural (*i.e.* applying at the level of surface representations), it will be listener-oriented (*i.e.* applying at the mappings between auditory cues and surface segments).

3.3 Syllabic status of the prepositions

Some early studies on Czech phonology have questioned the syllabic status of the non-syllabic prepositions. Kučera (1961:72) argues that they are "isolated consonantal microsegments […] [that] do not constitute syllables but are associated with the following syllable. […] and are best considered as special pre-syllabic segments".

I argue that the preposition is either fully aligned to the following onset, which happens if it is non-vocalized, or forms a syllable itself, which happens if it is vocalized. In other words, I will try to show that no other concept such as pre-syllabic segments, which lies somewhere in-between the full alignment and full syllabification, is necessary.

In Section 2 of the present study it was seen that in most cases, onset properties predict the form that the preposition takes. The prosodic structure of the whole resulting preposition-word sequence also affects the prepositional form to a certain degree – the preposition is either 'allowed' to vocalize and add an extra syllable to the prosodic word, or the preposition must be aligned to the onset of the following syllable to avoid the addition of an extra syllable to the word. The fact that the non-vocalized preposition does not seem to affect the syllabic pattern of the sequence is a strong argument supporting the idea that non-vocalized prepositions are fully aligned with the following onset. Another argument favoring full alignment can be seen in the preference for a vocalized preposition before a sonorant-obstruent onset. If the non-vocalized form of the preposition were used before a /SO/-initial word, it would either become an onset of a newly emerged syllable, in which a consonant changes its identity and becomes a vowel, or it would become a part of a disallowed cluster with respect to the sonority contour.[16] Full alignment of the preposition to the following onset is what most of the literature argues for as well (e.g. Palková 1994:271-2).

3.4 The historical perspective

Studies that elaborate in more detail on why in certain environments vocalized forms of either prepositions or prefixes occur refer in the first place to the historical development of Czech, more specifically to yers. A brief note on the basics of the yer theory follows; for more information about yers and the historical processes associated with them, the reader is referred to e.g. Mann (1977); for processes related to prefixes (and prepositions), see Scheer (1996), Ziková (2008), and references therein.

Yers (the front yer ь and the back yer ъ, corresponding to /ĭ/ and /ŭ/ respectively) were very short reduced vowels of Proto-Slavonic that disappeared or were vocalized; in Czech these were replaced by the vowel /ɛ/. In a word that contained one or more yers, those yers in weak positions (*i.e.* the odd yers when counting from the final syllabic position) were deleted. When more yers followed each other, those yers in strong positions (*i.e.* the even yers when counting yers from the final syllabic position) were vocalized (Dickins 1998). For instance, historically the meaning 'with a dog' was *sъ pьsъmь*, vocalizing the strong yers we arrive at *se pьsemь*; when then the weak yers are deleted we get the form which is attested in present-day Czech *se psem*. This part of the theory can thus explain vocalizations in *se psem* and *ke dni*, but would also predict forms such as **k komu* (*kъ komu → k komu*) or **se ptákem*

[16] Hála (1962:286) has suggested that the sonorant in the sonorant-obstruent word-initial clusters is a 'side syllable'. He points out that the controversial debate about these side syllables is only relevant to the diachronic study of sound change but not to the synchronic description of sound patterns. Since the aim of this study is to provide purely synchronic explanations, I do not consider 'side syllables' to be of any relevance here.

(*sъ pьtákьmь → se pьtákemь → se ptákem*).[17] Therefore, it has been suggested that a 'secondary vocalization' and 'devocalization' took place (Komárek 1962:149). The secondary vocalization 'inserted' an /ɛ/ between the preposition and the first consonant of the noun if they were identical (Dickins (1998) calls this a 'homorganic rule'); e.g. *k komu* became *ke komu*. The devocalization, on the other hand, deleted the prepositional /ɛ/ from forms such as *se ptákem* that became *s ptákem* (for a summary see Scheer 1996:107).

The question that immediately arises is whether proposing explanations of the vocalizations that we observe today by referring to yer deletions and yer vocalizations is necessary if these explanations alone do not account for most of the contemporary data. Moreover, it is seen that this theory, which assumes that vocalizations and deletions once either did or did not take place in particular environments, can by no means account for the high within- and between-speaker variability in prepositional vocalizations observed in present-day Czech.

In line with the observations and the facts reported in the literature and in several previous studies (e.g. Komárek 1962, Scheer 1996, Dickins 1998), I conclude that with the yer-theory alone we are not able to predict the current situation in the language. Additionally, regarding the historical yer-related processes in prepositions, the contemporary language data are not transparent anymore and cannot predict where the yers were deleted and where they were vocalized. Therefore, I argue that for explaining the prepositional vocalizations it is neither necessary nor beneficial to refer to yers or to any other diachronic processes. A crucial argument that I believe is worth stating here is that the language user has no access to what the structures of her language looked like in the past and how they developed diachronically.[18]

---

[17] This is also where the distinctive behavior of prepositions and prefixes may come from: prefixes have an origin in prepositions, but once they are attached to a word, and yer deletions and vocalizations have taken place, it is not possible to modify them further; they are a part of the word – as opposed to prepositions. That is, the yer theory alone seems applicable to prefixes (see Ziková 2008) but not to prepositions.

[18] There are also regular synchronic (and thus 'transparent') vowel-zero alternations in various word forms in Czech, e.g. *d**e**n* 'a day-Nom' and *dne* 'a day-Gen', or *přiš**e**l* 'he came' and *přišla* 'she came'. Importantly, since there is no variation at all in these words with respect to whether the ***e*** shows

4 The present analysis

4.1 The framework

In this study I create an Optimality Theoretic (OT) model of the prepositional vocalization in the framework of Parallel Bidirectional Phonology and Phonetics (BiPhon; Boersma 2007, 2009). The model is sketched in Figure (5) (adapted from Boersma 2009). Bidirectionality means that both the listener and the speaker use the same constraint rankings and that the constraints express mappings between various levels of representation in both directions. Parallelism means that the multiple levels of representation through which the comprehension and production run are evaluated in parallel, thus allowing constraints on different levels of representation to interact. As an evaluation strategy, I use the Stochastic version of OT (Boersma 1997). As opposed to the traditional version of OT developed and introduced by Prince & Smolensky (1993), in which constraints have fixed rankings, constraints in Stochastic OT are ranked along a continuous scale. During evaluation a random amount of noise is added to the ranking values of the constraints; the ranking value plus the noise then result in a disharmony value. The ordering of the disharmony values therefore does not necessarily correspond to the ordering of the ranking values (*i.e.* constraints with nearby ranking values may be 're-ranked' during some evaluations). Stochastic OT is therefore able to account for variation in the output.

Section 2 showed that the prepositional vocalization depends on many factors that interact with each other and is quite variable. For these two reasons, the BiPhon framework and Stochastic OT as the evaluation strategy are employed in the present study.

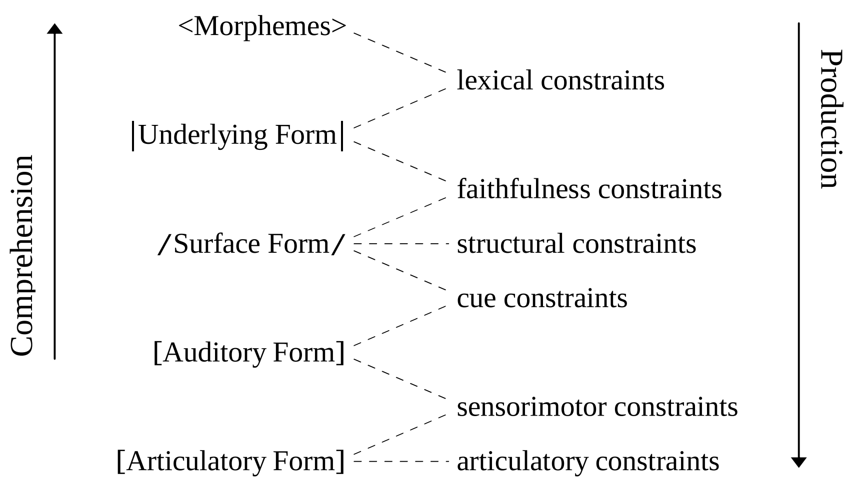
The model of Czech prepositional vocalization that I propose needs at least five levels of representation (which is also the minimal required number of levels for a phonological theory, according to Boersma 2009). As is seen in Figure (5), these five levels include a level at which the morpheme is specified, two levels of phonological representation, *i.e.* the underlying form and the surface form, and two levels of phonetic representation, *i.e.* the auditory form and the articulatory form. Four sets of constraints evaluate the relation between

---

up or not (contrary to the prepositional *e*), these paradigms may well be learnt. It is however not likely that the *prepositional* vocalizations, in which we do not observe any regularity but great variation, can be learnt as word paradigms.

each pair of neighboring levels of representation. The association between the morpheme and the underlying form is handled by lexical constraints, faithfulness constraints evaluate the similarity between the underlying form and the surface form, cue constraints express the mapping between the auditory form to the surface form, and the relation between the articulatory form and the auditory form is expressed in terms of sensorimotor constraints. Two sets of constraints operate at one level of representation alone: these are structural constraints, which evaluate the surface form, and articulatory constraints, which evaluate the articulatory form. For a more detailed description of the various levels of representation and constraints see Boersma (1998, 2007, 2009), Boersma and Hamann (to appear), and Apoussidou (2007). Figure (5) shows the BiPhon model with five levels of representation and the mappings between them.

Importantly, the model proposed in the present study is listener-oriented. Contrary to the majority of previous studies, which assumed that the vocalization occurs to lessen articulatory effort, I argue that the reason for the vocalized form to be used is to facilitate perception. More specifically, when speakers produce the vocalized forms, it is for the listener to be able to recover the morpheme (*i.e.* the preposition), which would not be possible if the non-vocalized form were used instead. This is elaborated on in more detail in the following sections. For creating the OT grammar, which is dealt with in this section, as well as for running the learning simulations, which are described in Section 5, I use the computer program Praat (Boersma & Weenink 2009).



(5) **Figure**: The five levels of representation and the constraints of the BiPhon model.

4.2 Simplified language data

The language data described in Section 2 will be largely simplified so that the whole analysis of the prepositional vocalizations can be done within the scope of the present study. Note however that the model created on the basis of this abstraction will still be able to predict correctly any of the prepositional forms that we can find in the target language. In Table (6) I give a list of the forms used as model examples of the real data, and a description of what part of the language data they stand for.

(6) Table: Left column: the real data of the target language split into categories on the basis of word length, onset complexity and onset similarity to the preposition. Right column: the words used in the model to represent the respective parts of reality.

real data	**model example**		
non-syllabic prepositions	k	<to>	/k/
vocalized prepositions	ke	<to>	/kɛ/
words of any length with simple onsets			
onset different from the preposition			
voiceless onset	týmu	<team-DAT>	/ti:mu/
voiced onset	domu	<house-DAT>	/domu/
onset identical to the preposition			
voiceless onset	kolu	<bike-DAT>	/kolu/
voiced onset	gólu	<goal-DAT>	/go:lu/
words with complex onsets			
at least bi-syllabic words			
resulting onset very similar	škvíře	<chink-DAT>	/ʃkvi:r̝ɛ/
resulting onset less similar	psovi	<dog2-DAT>[19]	/psovɪ/

[19] In the corpus there occur two forms that mean ‘a dog+Dat.’, *psovi* and *psu*. I mark the former one as <dog2> and the latter as <dog1>. Where the difference between these two forms lies

resulting onset not similar	plotu	<fence-DAT>	/plotu/
sonorant-obstruent onset	rtuti	<mercury-DAT>	/rtucɪ/
mono-syllabic words	psu	<dog1-DAT>	/psu/
vowel-initial words	autu	<car-DAT>	/ʔautu/

There are in Czech, naturally, also complex onsets starting in a voiced obstruent. For reasons of space I do not include these since it is explanatory enough to illustrate the effects related to voicing on the behavior of simple onsets; voiced complex onsets undergo the same voicing effects as the simple ones. Also, note that all the onset forms that were listed in Section 2, Table (3), are not represented individually but are divided into 3 classes according to the similarity within the onset, which results from linking the preposition and the word. These three classes are: onsets very similar to the preposition (these are often the most complex onsets as well), onsets of medium similarity, and onsets with hardly any similarity between its segments. The non-vocalized forms of the prepositions are represented here by the preposition *k* and the vocalized prepositions by its counterpart *ke*. The lexical constraints on the prepositional forms that I will employ in the present analysis are given in (7).

(7) Lexical constraints

*<to> |k| do not connect the morpheme <to> to the underlying form |k|

*<to> |kɛ| do not connect the morpheme <to> to the underlying form |kɛ|

Note that there must be a large number of other constraints such as *<to> |zu|, which are ranked very high since the form |zu| never occurs as a preposition with the meaning <to>, and |zu| is also unlikely to be the underlying form (UF) of /kɛ/ because it violates faithfulness in many respects (e.g. the features voicing, place, manner). I argue that the

---

and why there are two such allomorphs is not relevant for the present study. These two words were selected to illustrate the working of the present grammar on prosodically different words with one and the same onset.

constraint *<to> |g| is also high ranked; it is unlikely that the UF of the preposition *k* would be |g| (despite the fact that we do find forms such as /gdomu/, see Section 2) because we do not find prepositional forms such as /gɛ/. It then remains to be found whether the UF is |k| or |kɛ|, or whether there is allomorphy underlyingly, which is one of the aims of the present study and is dealt with in Section 5.

4.3 Simple onsets

In (8) and (9) I list the four faithfulness constraints and one structural constraint that are needed to account for the vocalization in some words, which I will illustrate below.

(8) Faithfulness constraints

MAX	a segment in the UF must have a corresponding segment in the surface form (SF)
DEP	a segment in the SF must have a corresponding segment in the UF
IDword	the voicing specification of segments in the UF and SF of a lexical word must be identical
IDprep	the voicing specification of segments in the UF and SF of a preposition must be identical

(9) Structural constraint

VOI	obstruents in a cluster have the same feature specification for voicing (*i.e.* voicing assimilation)

To clarify the functioning of the constraints that I have introduced, I will now show the phonological production of the words with simple onsets; this is done in Tableaus (10) and

(11). In the Tableaus, the mapping between the three forms (SF-AudF-ArtF) is ‘faithful’. That is, every segment and all its features in the SF are implemented auditorily, and the mapping between the ArtF and the AudF is perfect. Perfect mapping between the ArtF and the AudF means that the adult speaker has the knowledge of what articulatory gestures yield the auditory events that she wants to produce, *i.e.* this is a perfect sensorimotor knowledge (similarly to what Boersma (2009) assumes). Note that the notation used in this paper for the AudF and ArtF is a shorthand for the auditory and articulatory events respectively; for instance, the phoneme represented in the SF as /k/ would ideally correspond to [k] in the phonetic form, which is a shorthand for [k_^{k}]$_{\text{Aud}}$ (*i.e.* transitions to velar closure, voiceless silence, short velar burst) or for [back of the tongue moving towards and touching the velum, closure with no vocal fold vibration, release of the closure]$_{\text{Art}}$. In Tableaus, the AudF and the ArtF will be formally represented by one form only – the overt form (OF). Importantly, the distinction between the AudF and ArtF must be kept, because I will introduce different constraints, some of which will apply to the ArtF (articulatory constraints), while some will apply to the mapping between the AudF and the SF (cue constraints).

(10) Tableau: Phonological production of <to+team>.

<table>
<tr><th><to+team></th><th>MAX</th><th>DEP</th><th>*<to> |kɛ|</th><th>*<to> |k|</th></tr>
<tr><td>☞ |k+ti:mu| /.kti:.mu./ [kti:mu]</td><td></td><td></td><td></td><td>*</td></tr>
<tr><td>|k+ti:mu| /.kɛ.ti:.mu./ [kɛti:mu]</td><td></td><td>*!</td><td></td><td>*</td></tr>
<tr><td>|kɛ+ti:mu| /.kti:.mu./ [kti:mu]</td><td>*!</td><td></td><td>*</td><td></td></tr>
<tr><td>|kɛ+ti:mu| /.kɛ.ti:.mu./ [kɛti:mu]</td><td></td><td></td><td>*!</td><td></td></tr>
</table>

In Section 3, I noted that the literature seems to consider |k| the UF. Moreover, the corpus shows that the form /k/ is more than 5 times more frequent than /kɛ/. With the constraint ranking of Tableaus (10) and (11), the grammar will prefer candidates with |k| as the UF as well. The AudFs that speakers of Czech produce for the morphemes <to+team> and <to+house> are [ktomu] and [gdomu], respectively. In Tableau (10) there are two candidates that do not violate any faithfulness constraints, and two candidates that do violate faithfulness constraints; the two faithful candidates will thus be preferred (*i.e.* candidate 1 and 4). Since one of them (candidate 4) does not contain the OF that speakers produce, *i.e.* [kɛtiːmu], the winner must be the first candidate. To have the grammar select the first candidate as optimal, the constraint *<to> |kɛ| must be ranked above *<to> |k|. Similarly, in Tableau (11), there are only two candidates whose OF is [gdomu]. Since one of them violates MAX, the candidate in the second row will win. To have this candidate win, *<to> |kɛ| is ranked above *<to> |k| to punish the candidate in the last row, which contains an OF that Czech speakers do not produce, *i.e.* [kɛdomu].

In this section, I thus assume that the constraint *<to> |kɛ| is ranked high. However, it might be the case that real Czech speakers (whose grammar contains many more constraints than the four of Tableau (10)) prefer |kɛ| as the UF, or have two allomorphs underlyingly (*i.e.* both |k| and |kɛ|). Learning simulations, which will be presented in Section 5, will determine whether the UF is |k| or |kɛ|, or both.

(11) Tableau: Phonological production of <to+house>.

<to+house>	*<to> |kɛ|	MAX	DEP	ID word	Voi	*<to> |k|	ID prep
|k+domu| /.kdo.mu./ [kdomu]					*!	*	
☞ |k+domu| /.gdo.mu./[gdomu]						*	*
|k+domu| /.kto.mu./ [ktomu]				*!		*	
|k+domu| /.kɛ.do.mu./ [kɛdomu]			*!			*	
|kɛ+domu| /.kdo.mu./ [kdomu]	*!	*			*		
|kɛ+domu| /.gdo.mu./ [gdomu]	*!	*					*
|kɛ+domu| /.kto.mu./ [ktomu]	*!	*		*			
|kɛ+domu| /.kɛdo.mu./ [kɛdomu]	*!						

As becomes apparent in (11), the other crucial rankings are as follows: IDprep has to be outranked by both IDword (to rule out forms such as /.kto.mu./) and by VOI (to rule out forms such as /.kdo.mu./).

Thus far, I have assumed a ‘faithful-like’ mapping between the SF and the OF. Because the aim of this study is to provide a model of a grammar that will account for mappings between all the five levels of Figure (5) (*i.e.* not only <morpheme> ↔ UF ↔ SF but also SF ↔ OF), I drop this assumption and introduce cue constraints that handle the SF ↔ OF mapping; this is done in (12).

(12) Cue constraints

*/-,+/ [+,-]	the presence of periodicity does not correspond to a voiceless segment in the SF, and the absence of periodicity does not correspond to a voiced segment in the SF
*/ / [x]	the presence of auditory events does not correspond to the absence of a segment in the SF
*/x/ [ ]	the absence of auditory events does not correspond to a segment in the SF

The first cue constraint in (12) punishes the mapping of a phonologically voiced segment to a sound without vocal fold vibration (*i.e.* periodic sound) and vice versa. It thus eliminates perceptions such as [gdomu] → /.kdo.mu./ or [ktomu] → /.gdo.mu./, and phonetic implementations such as /.kdo.mu./ → [gdomu] or /.gdo.mu./ → [ktomu].

The constraints */ / [x] and */x/ [ ] are rather intuitive, already for a child learner, because if a speech sound is auditorily present, it should be mapped onto *something*, and vice versa. These two constraints must naturally be included in the present model as well, because the model deals with a potential vowel insertion (or vowel deletion). The role of these constraints will become clear later on in this study in Tableaus (15), (18) and (19).

4.4 Identical onsets

This section will show how the grammar accounts for the obligatory vocalization in prepositions that are followed by onsets starting with a consonant that is identical to the preposition.

To have the grammar produce correctly words that have onsets that are identical to the preposition, I will use the articulatory constraint in (13) and the cue constraint in (14).

(13) Articulatory constraint

*$[C_iC_i]_{Art}$	do not produce two adjacent identical separate consonantal articulatory gestures

(14) Cue constraint

*/CC/ $[\_C:]_{Aud}$	an auditorily (prolonged) single consonant that follows a pause does not correspond to two consonantal segments in the SF (Boersma 1998 introduced a similar constraint)

I showed in Tableau (11) that the constraint VOI is ranked high. In Section 4.3 I introduced the voicing cue constraint */-,+/ [+,-] and explained how it works. Because of these two high-ranked constraints, adjacent obstruents that are underlyingly identical in place and manner but not in voicing will turn out to have the same specification for voicing in the SF and will thus be absolutely identical in the OF as well.[20]

The Czech underlying phoneme inventory does not contain long consonants at the surface. Adjacent identical consonants are usually pronounced as a single (slightly prolonged)

[20] A recent study presented an analysis of Polish non-syllabic proclitics and their vocalizations (Pająk & Bakovic 2009); they showed that since adjacent obstruents have to agree in voicing on the surface, a structural constraint that prevents geminates applies and forces vowel insertion.

consonant.[21] Therefore, in (13) I propose the articulatory constraint $[C_iC_i]_{Art}$. This constraint militates against two adjacent identical articulatory gestures and would therefore indirectly favor that two adjacent identical segments be phonetically implemented as a single (prolonged) segment instead. Clearly, as I already speculated in Section 3.2, a single (even though prolonged) consonantal articulation requires less effort than two full adjacent consonantal gestures.

Importantly, there is also the cue constraint */CC/ $[\_C\text{ː}]_{Aud}$; introduced in (14). This constraint does not favor mapping a single (even though prolonged) post-pausal consonantal segment onto two consonants in the SF. The notion of a pause in the formulation of this constraint reflects the fact that after a pause the auditory cues to a consonant are reduced substantially. If the sequence of two identical Cs is onset-initial with no vocalic segment preceding, formant transitions *into* the initial consonant are missing. Therefore it is hard to recover two surface segments in the complete absence of some consonantal cues. On the other hand, if the OF is $[VC\text{ː}V]_{Aud}$, there are formant transitions *into* the [Cː], which are present in the [V] that precedes, (and clearly also transitions *from* the [Cː], which are present in the [V] that follows). Also, in $[VC\text{ː}V]_{Aud}$, thanks to the preceding vowel, the onset of the stop closure is very well audible, unlike when the [Cː] is post-pausal.

The perfect sensorimotor knowledge that I described in Section 4.3 means that the correspondence between the AudF and ArtF is perfect. Therefore, the Tableaus do not contain separate notations for both the AudF and ArtF but only one OF. Nevertheless, as I pointed out in Section 4.3 it is still necessary to distinguish between the two levels of representation as the grammar now contains constraints that evaluate the ArtF and constraints that evaluate the mapping between the AudF and SF. These two sets of constraints are not interchangeable. Tableau (15) shows how the grammar with these newly introduced cue and articulatory constraints handles the identical onsets in production.

It is seen that only those candidates that contain the OF [kɛkolu] do not militate against either of the two new constraints. Recall that as shown in Tableau (10), most plausibly

[21] And as two separate consonants if the ‘single’ pronunciation would result in confusion (e.g. *bez země* ‘without ground’ and *beze mě* ‘without me’); see Hála (1962:216) and Palková (1994:327-8).

the constraint *<to> |kɛ| is ranked high. Note that the phonetic form of the meaning <to+bike> is [kɛkolu].

(15) Tableau: Production of <to+bike>.

<table>
<tr><th><to+bike></th><th>*<to> |kɛ|</th><th>$*[C_iC_i]_{Art}$</th><th>*/CC/ [_C:]</th><th>DEP</th><th>*/ / [x]</th><th>*<to> |k|</th><th>IDprep</th></tr>
<tr><td>|k+kolu| /.kko.lu./ [kkolu]</td><td></td><td>*!</td><td></td><td></td><td></td><td>*</td><td></td></tr>
<tr><td>|k+kolu| /.kko.lu./ [k:olu]</td><td></td><td></td><td>*!</td><td></td><td></td><td>*</td><td></td></tr>
<tr><td>☞|k+kolu| /.kko.lu./ [kɛkolu]</td><td></td><td></td><td></td><td></td><td>*</td><td>*</td><td></td></tr>
<tr><td>☞|k+kolu| /.kɛ.ko.lu./ [kɛkolu]</td><td></td><td></td><td></td><td>*</td><td></td><td>*</td><td></td></tr>
<tr><td>|kɛ+kolu| /.kɛ.ko.lu./ [kɛkolu]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

In Tableau (15) candidates 3 and 4 have the UF |k+kolu| and the OF [kɛkolu]. In order to have one of them win, it is crucial that the constraints $*[C_iC_i]_{Art}$, */CC/ [_C:] both outrank DEP or/and */ / [x]. It is clear that if DEP were ranked above */ / [x], then [ɛ] would be inserted in phonetic implementation; if on the other hand */ / [x] were ranked above DEP, /ɛ/ would be inserted during phonological production. The next section discusses which of these two analyses is appropriate.

4.5 Phonological vs. phonetic ɛ-insertion

To see which of the two winning candidates in (15) is more plausible for the Czech speaker, we have to consider one crucial characteristic of Czech, which was already noted in Section 2. Stress in Czech falls on the first syllable of a prosodic word. If the preposition is syllabic, it carries the main stress. In other words, the [**ɛ**] in [k**ɛ**kolu] has the phonetic properties of a stressed vowel (intensity, pitch[22]). I use a bold letter to symbolize phonetic prominence in the

[22] In Czech, stressed vowels are not lengthened phonetically (Palková 1994:279).

auditory form (*i.e.* the phonetic properties of stress). The syllabic and the prosodic structure are reflected in the SF. The stressed vowel has to be represented in the SF as the first vowel of a prosodic word. Two constraints handle stress assignment, a cue constraint (16), and a structural constraint (17).

(16) Cue constraint

*/a, á/ [**a**, a]	a phonetically prominent segment [**a**] does not correspond to an unstressed vowel /a/; a phonetically non-prominent segment [a] does not correspond to a stressed vowel /á/

(17) Structural constraint

STRESS 1^{st}	the first syllable of a prosodic word is stressed

The structural constraint in (17) could be rewritten with TROCHAIC, *i.e.* feet are trochaic, and ALL-FEET-LEFT (McCarthy & Prince 1993). For reasons of simplicity, I do not mark foot structure in the SF of the candidates and accordingly prefer to use the formulation in (17). Similarly, prosodic-word boundaries are not marked, because the edges of the SF of all candidates in this paper are the edges of one prosodic word. Both constraints are high-ranked; the high-ranking of */a, á/ [**a**, a] is valid cross-linguistically (although the definition of 'prominent' is language-specific), and the constraint STRESS 1^{st} is undominated in a word-initial-stress language such as Czech.

Tableau (18) illustrates how the two high-ranked 'stress constraints' work; the Tableau also shows that eventually the cue constraint */ / [x] has to be ranked above DEP to rule out incorrect candidates such as /.kkó.lu./ [kεk**O**lu] that do not violate any of the 'stress constraints' but that yield auditory forms that Czech speakers do not produce.

(18) Tableau: Production of <to+bike> with the constraints for stress.

<to+bike>	*/a, á/ [**a**, a]	STRESS 1^{st}	*/ / [x]	*/x/ [ ]	DEP
|k+kolu| /.kkó.lu./ [k**ε**kolu]	*!		*		

|k+kolu| /.kkó.lu./ [kɛk**o**lu]			*!		
|k+kolu| /.kko.lu./ [k**ɛ**kolu]		*!	*		
|k+kolu| /.kɛ́.ko.lu./ [kɛk**o**lu]	*!				
☞ |k+kolu| /.kɛ́.ko.lu./ [k**ɛ**kolu]					*
|k+kolu| /.kɛ́.ko.lu./ [kɛk**o**lu]	*!				

In Tableau (18) it is seen that neither candidate 1 nor candidate 3 can win, despite the fact that their OF is what the real speakers produce. To prevent candidate 2 from winning – because its OF is not the form that Czech speakers produce – */ / [x] has to outrank DEP. This is because of the language-specific stress assignment: if vocalized, the preposition is stressed. Stress has to be represented in the SF, and in Czech the stressed vowel has to be the nucleus of the first syllable of a word. I conclude that the insertion of /ɛ/ happens in the phonology (*i.e.* in the SF) and not in the phonetics.

In (19) I show how production would fail if */ / [x] were ranked below DEP. (The symbol ☹ denotes the candidate that should have won since it contains the OF that real speakers produce and does not violate either of the two undominated 'stress constraints'; the symbol ☞ denotes the winner according to the Tableau.)

(19) Tableau: Failed production of <k+kolu> if DEP outranks */ / [x].

<to+bike>	*/a, á/ [**a**, a]	STRESS 1st	DEP	*/ / [x]	*/x/ [ ]
|k+kolu| /.kkó.lu./ [k**ɛ**kolu]	*!			*	
☞ |k+kolu| /.kkó.lu./ [kɛk**o**lu]				*	
|k+kolu| /.kko.lu./ [k**ɛ**kolu]		*!		*	
|k+kolu| /.kɛ́.ko.lu./ [kɛk**o**lu]	*!				
☹ |k+kolu| /.kɛ́.ko.lu./ [k**ɛ**kolu]			*!		
|k+kolu| /.kɛ́.ko.lu./ [kɛk**o**lu]	*!				

4.6 Listener-orientedness

As has been already put forth, the present model is listener-oriented. Although I have been presenting the workings of constraints and their rankings in production tableaus, it is important to note that it were the cue constraints, *i.e.* the mappings between the SF and the AudF, that punished candidates like <to+bike> |k+kolu| /.kko.lu./ [k:olu].

We have seen that because articulatory constraints do not favor the form [kkolu] speakers would tend to produce [k:olu] for the SF /.kko.lu./. Indeed, merging two identical adjacent consonants into one prolonged consonant is what Czech speakers usually do (Palková 1994:327-8; Hála 1962:216). It is thus the cue constraint */CC/ [_C:] that ensures that in the case of word onsets identical to the preposition, the preposition is always vocalized. In other words, the high ranking of the cue constraint */CC/ [_C:] is there for the listener to be able to recover the message.[23] Tableau (20) shows that comprehension fails if the preposition is not vocalized, and Tableau (21) shows that the vocalized form is perceived correctly.

The Tableaus (20) and (21) are comprehension tableaus, of the type described by Boersma (2009). Unlike in traditional OT production tableaus, the input in these comprehension tableaus is the OF; the candidates are triplets SF-UF-<morpheme>.

---

[23] In support of my analysis with the articulatory constraint $*[C_iC_i]_{Art}$ and the cue constraint */CC/ [_C:] is the fact that Czech speakers make mistakes in writing in the use of the preposition in the only example of identical word onset, in which vocalization is not realized, *s sebou* 'with oneself' (see Dickins (1998), who also points out that this is the only case of identity in Czech in which the preposition remained non-vocalized). The speakers of Czech very often mistakenly write *sebou* for the meaning 'with oneself' (while *sebou* means 'oneself-INSTR'). According to my analysis, this is because the absence of vocalization in *s sebou* creates an environment for the articulatory constraint $*[C_iC_i]_{Art}$ that does not favor pronouncing two adjacent identical consonants as two separate gestures. Speakers thus realize the sequence *s sebou* as [s:ɛbou] or [sˑɛbou] or even [sɛbou] (instead of [ssɛbou]). When the listener hears the OF [s:ɛbou], the cue constraint */CC/ [_C:] prevents her from perceiving two consonantal segments in the SF. The listener thus maps the form [s:ɛbou] onto /.sɛ.bou./ (and not onto /.ssɛ.bou./). Having lost the preposition during comprehension, when asked to replicate what she had heard (e.g. in writing), the language user will produce (e.g. write) *sebou* but not *s sebou*.

Importantly, because the model is bidirectional, the constraints and rankings are identical to those used in the production tableaus above.[24]

(20) Tableau: Comprehension of [k:olu] as <coke+Acc.>.[25]

[k:olu]	*<to> | |	MAX	$*[C_iC_i]_{Art}$	*/CC/ [_C:]	*[x] / /	*[ ] /x/	DEP	*<to> |k|	*<coke> |kolu|	*<a bike> |kolu|
☞ /.ko.lu./ |kolu|<coke+Acc.>									*	
/.ko.lu./ |kolu| <to+bike>	*!									*
/.ko.lu./ |k+kolu| <to+bike>		*!						*		*
/.kɛ.ko.lu./ |k+kolu| <bike>						*!	*	*		*
/.kko.lu./ |k+kolu| <to+bike>				*!				*		*

(21) Tableau: Comprehension of [kɛkolu] as <to+bike>.

[kɛkolu]	*< > | x |	MAX	$*[C_iC_i]_{Art}$	*/CC/ [_C:]	*[x] / /	*[ ] /x/	DEP	*<to> |k|	*<coke> |kolu|	*<a bike> |kolu|
/.ko.lu./ |kolu|<coke+Acc.>					*!*				*	
/.ko.lu./ |kolu| <to+bike>		*!			*			*		*
/.ko.lu./ |k+kolu| <to+bike>					*!				*	
/.kɛ.ko.lu./ |k+kolu| <to+bike>	*!						*	*		*
☞/.kɛ.ko.lu./ |k+kolu| <to+bike>							*	*		*

[24] Tableaus (20) and (21) contain high-ranked lexical constraints *<x> | |, and *< > |x| that militate against connecting a morpheme with an empty UF, and against connecting an empty morpheme with an underlying segment, respectively.

[25] The OF that is in Czech the realization of the morpheme <coke+Acc> is [kolu].

4.7 Complex onsets

In Section 4.4, a cue constraint was introduced that favored /ɛ/ insertion. Initially, it was seen that Czech speakers do not like articulating adjacent identical consonantal gestures separately, and they often collapse /CC/ into [C:] or even [C]. Then, I showed that there is the cue constraint */CC/ [_C:], which militates against perceiving two adjacent identical segments word-initially (*i.e.* after a pause).

Such analysis thus has a similar effect as the Obligatory Contour Principle (OCP), which forbids surface representations with two identical adjacent segments. The OCP was first proposed in tonal phonology by Goldsmith (1976) and was later extended to segmental phonology (e.g. McCarthy (1986)). One process that the OCP triggers is vowel epenthesis between identical segments. In the present model, this OCP-like effect is achieved by the cue constraint (taking into account that an articulatory constraint militating against two adjacent identical articulations creates an environment for the cue constraint to operate). OCP-like processes without a reference to the structural OCP itself but to auditory cues instead have been introduced by Boersma (1998, 2000).

The idea of cue constraints accounting for OCP-like effects is developed further in the present section. The constraints that I will introduce now are based on the constraint */CC/ [_C:], which handles the linking of prepositions to simple onsets. In this section, words beginning in complex onsets are dealt with. The consonant cluster, which is a result of aligning the non-vocalized non-syllabic preposition to the word, contains thus three or more segments. Recall that all Czech words that begin in complex clusters containing more than two segments are represented by a CCC-initial word in the present study (/ʃkviːr̝̊ɛ/). If the non-vocalized form of the preposition is used then CC clusters (here represented by /plotu/ and /psovɪ/) become CCC (*i.e.* /kplotu/ and /kpsovɪ/), and the CCC clusters (*i.e.* /ʃkviːr̝̊ɛ/) become CCCC (*i.e.* /kʃkviːr̝̊ɛ/). Section 2 showed that the similarity in the resulting onset cluster contributes to the distribution of vocalized forms, and that very often a greater similarity relates to a greater complexity of the onsets at the same time. I introduce constraints that are meant to account for the similarity within the resulting cluster. It will be seen that the constraints that I present in (22) do not only punish similarity but importantly, an effect that emerges is that they punish complexity of the onset as well.

(22) Cue constraints

*/CCCC/ [7cue]	7 different consonantal cues do not correspond to 4 consonantal segments in the SF
*/CCC/ [6cue]	6 different consonantal cues do not correspond to 3 consonantal segments in the SF
*/CCC/ [7cue]	7 different consonantal cues do not correspond to 3 consonantal segments in the SF

Only three constraints are presented in (22), because these are relevant for the words selected in this study to represent the real data. However, there are many more similar constraints that evaluate the mapping between any number of cues and any number of segments. The mechanism that stands behind the formulation of these constraints and behind their rankings relative to each other is described in the following paragraph. Importantly, it has to be noted that this mechanism is oversimplified; more complexity and detail would have to be introduced to achieve a realistic analysis. Despite this apparent simplification, I argue that the actual mechanism of how in reality auditory cues determine the prepositional vocalization is based on the idea that I present here.

I assume that a consonant is specified by three main distinctive auditory cues, which roughly correspond to three sets of distinctive features: (i) [the amount of noise, silence, or continuity]$_{Aud}$ corresponds to /manner/, (ii) [the formant(s)]$_{Aud}$ corresponds to /place/, (iii) [periodicity]$_{Aud}$ corresponds to /voicing/.[26] This idea that phonological features are primarily defined by auditory representation has a tradition that started with Jakobson *et al*. (1952), and is somewhat different from the claim that features are defined solely in articulatory terms

[26] The voicing feature is included here, because it is not only obstruents that can be a part of the cluster, but also sonorants, which are always voiced and do not trigger voicing assimilation. Therefore, consonants (*i.e.* both obstruents and sonorants) in a complex onset do not necessarily have to be identical in voicing, and these cue constraints credit this dissimilarity. This is why a cluster composed of both obstruents and sonorants, such as /ktl/, is considered less similar than a cluster where all Cs are obstruents, such as /kts/; even though all of /t/, /l/, and /s/ are alveolars, the cluster /ktl/ contains more cues thanks to the inclusion of voicing as a differentiating cue.

(Chomsky & Halle 1968). While the features manner and place are unary (*i.e.* any C can have a different manner and place feature from any other C, at least in a cluster of five Cs), the voicing feature is binary (*i.e.* the maximum number of different cues for voicing in a complex cluster of any size is 2).

To arrive at the maximum number of different cues specifying a cluster of a particular size, we multiply the number of consonants by 2 (because each C is specified by the auditory cues for place and manner), and add 2 to the product (because regarding voicing, a cluster of any size may contain maximally 2 different cues). A cluster of 3 consonantal segments is therefore expressed maximally by 8 different cues, and a cluster of 4 consonants by 10 different cues.

In the language, however, clusters often contain segments in which some features are identical to the features of other segments in the cluster. For instance, the cluster /pl/ contains the following features, *i.e.* auditory cues in its phonetic form [kpl]:[27] plosive + velar + plosive + bilabial + approximant + alveolar + (voiceless + voiced) = 7 *different* cues. The sequence [ksf] contains the cues plosive + velar + fricative + alveolar + fricative + labio-dental + (voiceless), only 6 of which are different. Since the maximum number of cues for a CCC cluster is 8, the loss of cues in /ksf/ is 8 – 6 = 2; similarly, the loss of cues in /kpl/ is 8 – 7 = 1. It was seen in Section 2 that clusters such as /sf/ occur with the vocalized preposition *ke* much more often than clusters such as /pl/. Unsurprisingly with the analysis just presented, but crucially for the Czech grammar, the loss of 2 cues is worse than the loss of 1 cue. Therefore, the SF of a CCC cluster prefers to be connected with an auditory form that loses 1 cue, rather than with a form that loses 2 cues; thus the ranking of the three new cue constraints is */CCCC/ [7cue] >> */CCC/ [6cue] >> */CCC/ [7cue], *i.e.* loss of 3 cues >> loss of 2 cues >> loss of 1 cue.

It is now apparent that the effect of these cue constraints and the way they are ordered is OCP-like and is listener-oriented. The more similarity there is among the consonants in the resulting preposition-word onset cluster, the more disfavored such a cluster is. In other words,

[27] Note that what I refer to as ‘plosive’ is, auditorily, a shorthand for ‘a silence followed by a burst auditorily’, ‘fricative’ means ‘the presence of fricative noise’, ‘bilabial’ would be ‘the presence of a lower formant’ (as compared to alveolar for example), etc.

the fewer cues the listener is given, the more difficulty she will have with perceiving the presence of all the segments. The variation in vocalizations with complex onsets that was described in Section 2 is summarized as follows: most complex and most similar onsets occur almost always (*i.e.* approximately in 90% of the cases) with vocalized prepositions, less complex and less similar onsets occur approximately equally often (*i.e.* approximately in 50% of the cases) with vocalized as with non-vocalized prepositions, and less complex and the least similar onsets occur almost always (*i.e.* approximately in 90% of the cases) with non-vocalized prepositions. Such variation can well be accounted for by Stochastic OT. In Tableaus (23) through (25) it is seen that if the three cue constraints and DEP are given ranking values that are removed from each other by only small distances on the ranking scale, Stochastic OT as an evaluation strategy will yield the desired variation. The Tableaus not only show how the given constraint rankings result in a variable output that contains both the vocalized and non-vocalized prepositional forms, but more importantly the output that these rankings yield is comparable to reality. Despite the fact that the real percentages may be different from the percentages yielded by this model, the tendency is undoubtedly in the same direction that is observed in the real language – more similar (and more complex) onsets do occur with vocalized prepositions more often than less similar (and less complex) onsets; see Section 2.2.

(23) Tableau: Production of <to+fence>.

<table>
<tr><td>ranking value</td><td>100</td><td>100</td><td>82</td><td>80.1</td><td>80</td><td>78</td><td rowspan="2">frequency of this winner</td></tr>
<tr><td><to+fence></td><td>*/ / [x]</td><td>*/x/ [ ]</td><td>*/CCCC/ [7cue]</td><td>*/CCC/ [6cue]</td><td>DEP</td><td>*/CCC/ [7cue]</td></tr>
<tr><td>☞ |k+plotu| /.kplo.tu./ [kplotu]</td><td></td><td></td><td></td><td></td><td></td><td>*</td><td>77%</td></tr>
<tr><td>|k+plotu| /.kplo.tu./ [kɛplotu]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>|k+plotu| /.kɛ.plo.tu./ [kɛplotu]</td><td></td><td></td><td></td><td></td><td>*!</td><td></td><td>23%</td></tr>
<tr><td>|k+plotu| /.kɛ.plo.tu./ [kplotu]</td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

(24) Tableau: Production of <to+dog>.

<table>
<tr><td>ranking value</td><td>100</td><td>100</td><td>82</td><td>80.1</td><td>80</td><td>78</td><td rowspan="2">frequency of this winner</td></tr>
<tr><td><to+dog2></td><td>*/ / [x]</td><td>*/x/ [ ]</td><td>*/CCCC/ [7cue]</td><td>*/CCC/ [6cue]</td><td>DEP</td><td>*/CCC/ [7cue]</td></tr>
<tr><td>|k+psovi| /.kpso.vi./ [kpsovi]</td><td></td><td></td><td></td><td>*!</td><td></td><td></td><td>48%</td></tr>
<tr><td>|k+psovi| /.kpso.vi./ [kɛpsovi]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>☞ |k+psovi| /.kɛ.pso.vi./ [kɛpsovi]</td><td></td><td></td><td></td><td></td><td>*</td><td></td><td>52%</td></tr>
<tr><td>|k+psovi| /.kɛ.pso.vi./ [kpsovi]</td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

(25) Tableau: Production of <to+chink>.

<table>
<tr><td>ranking value</td><td>100</td><td>100</td><td>82</td><td>80.1</td><td>80</td><td>78</td><td rowspan="2">frequency of this winner</td></tr>
<tr><td><to+chink></td><td>*/ / [x]</td><td>*/x/ [ ]</td><td>*/CCCC/ [7cue]</td><td>*/CCC/ [6cue]</td><td>DEP</td><td>*/CCC/ [7cue]</td></tr>
<tr><td>|k+ʃkviːr̝ɛ| /.kʃkviː.r̝ɛ./[kʃkviːr̝ɛ]</td><td></td><td></td><td>*!</td><td></td><td></td><td></td><td>19%</td></tr>
<tr><td>|k+ʃkviːr̝ɛ| /.kʃkviː.r̝ɛ./[kɛʃkviːr̝ɛ]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>☞ |k+ʃkviːr̝ɛ| /.kɛ.ʃkviː.r̝ɛ./[kɛʃkviːr̝ɛ]</td><td></td><td></td><td></td><td></td><td>*</td><td></td><td>81%</td></tr>
<tr><td>|k+ʃkviːr̝ɛ| /.kɛ.ʃkviː.r̝ɛ./[kʃkviːr̝ɛ]</td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

It is seen in Tableaus (23), (24) and (25) that the order of onsets starting from those that require vocalized preposition least often and ending with those onsets that require it most often is comparable to the tendencies in the corpus (described in Section 2). I conclude that it is the cue constraints whose interaction with DEP yields the observed variation in output. The present analysis also provides an argument against articulatory constraints being the cause of the vocalizations. Articulatory constraints militate against articulatory effort, and prefer as effortless articulations as possible (‘laziness constraints’; Kirchner 1998, Boersma 1998). Technically, every consonantal cue that is present auditorily has to be a result of some consonantal articulatory gesture. Therefore articulatory constraints would rank the forms that require more consonantal cues (*i.e.* more consonantal articulatory gestures within a short time) above articulations with fewer cues. As Tableau (26) shows, such an analysis would fail to yield the observed variation. Namely, in Tableau (26) the least similar onsets such as /kpl/, which in reality almost always occur without the vocalization, would – according to the grammar in which articulatory constraints interact with DEP – prefer the vocalized form more often than more similar onsets such as /kps/, which in reality contain the vocalization more often than /kpl/.[28]

(26) Tableau: Unattested variation in production with articulatory constraints.

<table>
<tr><td>ranking value</td><td>84</td><td>83</td><td>82</td><td>81</td><td>80</td><td>80</td><td rowspan="4">frequency of this winner</td></tr>
<tr><td><to+fence></td><td rowspan="3">*[10 cues] $_{Art}$</td><td rowspan="3">*[9 cues] $_{Art}$</td><td rowspan="3">*[8 cues] $_{Art}$</td><td rowspan="3">*[7 cues] $_{Art}$</td><td rowspan="3">*[6 cues] $_{Art}$</td><td rowspan="3">DEP</td></tr>
<tr><td><to+dog2></td></tr>
<tr><td><to+chink></td></tr>
<tr><td>☹ |k+plotu| /.kplo.tu./ [kplotu]</td><td></td><td></td><td></td><td>!*</td><td></td><td></td><td>30%</td></tr>
<tr><td>☞ |k+plotu| /.kɛ.plo.tu./ [kɛplotu]</td><td></td><td></td><td></td><td></td><td></td><td>*</td><td>70%</td></tr>
<tr><td>|k+psovi| /.kpso.vi./ [kpsovi]</td><td></td><td></td><td></td><td></td><td>*!</td><td></td><td>50%</td></tr>
<tr><td>☞ |k+psovi| /.kɛ.pso.vi./ [kɛpsovi]</td><td></td><td></td><td></td><td></td><td></td><td>*</td><td>50%</td></tr>
<tr><td>|k+ʃkviːr̪ɛ| /.kʃkviː.r̪ɛ./ [kʃkviːr̪ɛ]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td><td>10%</td></tr>
<tr><td>☞ |k+ʃkviːr̪ɛ| /.kɛ.ʃkviː.r̪ɛ./ [kɛ.ʃkviːr̪ɛ]</td><td></td><td></td><td></td><td></td><td></td><td>*</td><td>90%</td></tr>
</table>

The analysis of Tableau (26) has shown that a grammar with articulatory constraints (instead of cue constraints) cannot account for the variation that we observe in reality.

[28] In this alternative analysis with articulatory constraints I count how many different articulatory gestures have to be made in the sequence of the consonants, e.g. /kpl/ corresponds to voiceless + plosive + velar + bilabial + voiced + approximant + alveolar = 7 and /kps/ corresponds to voiceless + plosive + velar + bilabial + fricative + alveolar = 6.

4.8 Syllabic analysis

In Section 3, I noted that some previous studies have questioned the syllabic status of the non-vocalized preposition. Some have regarded them as only loosely associated to the following onset and called them 'pre-syllables' (Kučera 1961), while others have recognized them as being fully aligned with the following onset (Rubach 2000). This study aims at finding out which syllabic analysis is most plausible. Therefore, in (27) I introduce three structural constraints that handle syllabification.

(27) Structural constraints

ALIGN	syllable boundaries correspond to morpheme boundaries
PARSE	segments must be parsed into syllables
*O/N	an obstruent cannot be a syllable nucleus

This model thus aims at solving this syllabification issue, and therefore the constraints ALIGN, PARSE and *O/N are used. Generally in Czech, syllable boundaries coincide with morpheme boundaries (Palková 1994:271), hence the constraint ALIGN (after McCarthy & Prince 1993). Cross-linguistically, syllables typically require a vocalic nucleus and segments are parsed into syllables, hence the constraints * O/N (after Prince & Smolensky 1993) and PARSE (after PARSE by Tesar & Smolensky 2000, but operating at a lower prosodic level). If PARSE were ranked low, a possible analysis would be to keep the non-vocalized preposition extra-syllabic (*i.e.* 'pre-syllabic'), noted here as `/(k).ko.lu./`. If *O/N were ranked low, the non-vocalized preposition might form a syllable by itself, *i.e.* `/.k.ko.lu./`. Finally, if ALIGN were ranked low, the preposition would be a part of the following onset, *i.e.* `/.kko.lu./`, which, I will show, is a more plausible analysis than the former two. Tableau (28) shows how these three constraints work in the grammar, taking <to+bike> as the input to production.

(28) Tableau: Production of <to+bike>.

ranking value	100	100	100	100	100	100	100	80	50	20
<to+bike>	*/a, á/ [**a**, a]	STRESS 1st	*[C_iC_i]$_{Art}$	*/CC/ [_C:]	PARSE	*O/N	*/ / [x]	DEP	ALIGN	*<to> |k|
|k+kolu| /.kkó.lu./ [kk**O**lu]			*!						**	*
|k+kolu| /.kkó.lu./ [k:**O**lu]				*!					**	*
|k+kolu| /(k).kó.lu./ [kk**O**lu]			*!		*					*
|k+kolu| /(k).kó.lu./ [kεk**O**lu]					*!		*			*
|k+kolu| /.ḱ.ko.lu./ [**k**kolu]			*!			*				*
|k+kolu| /.ḱ.ko.lu./ [**k**εkolu]						*!	*			*
|k+kolu| /.kkó.lu./ [kεk**O**lu]							*!		**	*
☞|k+kolu| /.kέ.ko.lu./ [k**ε**kolu]								*	*	*

In Tableau (28), it is seen that even in the winning vocalized form, ALIGN is violated. It is therefore likely that ALIGN is ranked low. Section 5 will show whether the ranking that virtual learners arrive at resembles this assumption.

4.9 Monosyllables with complex onsets

In words with complex onsets, word length was seen to be an important factor contributing to the occurrence of vocalized prepositions (Section 2.2, examples in (4)). Recall that those monosyllabic words in which the monosyllable is light (e.g. *ke psu* ‘to a dog’), always require vocalized prepositions, while polysyllabic words with the same onset do not (e.g. *k psovi* ~ *ke psovi* ‘to a dog’). There also seems to be a tendency of heavy monosyllables to require the vocalized preposition less often than light monosyllables (*ke psům* and occasionally *k psům* ‘to dogs’, but always *ke psu*). To account for these observations I provide the model with two new structural constraints (29).

(29) Structural constraints

*FEETUN feet are not monosyllabic

MINWORD a light monosyllable does not constitute a prosodic word

Both constraints in (29) are high-ranked. *FEETUN (together with the high-ranked structural constraint STRESS 1^{st} introduced earlier) is a constraint that favors both of the stress patterns that are most frequent in Czech, *i.e.* a disyllabic trochee and a trisyllabic dactyl.[29] The notion of linking minimal word restrictions with foot structure was introduced by McCarthy & Prince (1986); Prince & Smolensky (1993) brought the minimal word restriction as a constraint into OT. The constraint reflects the language-specific requirement for what can be a minimal prosodic word. For Czech, if a prosodic word is monosyllabic, it has to be a heavy syllable, *i.e.* at least bimoraic (Hayes 1995:102); recall (footnote 15) that a coda counts as a mora and a long vowel counts as two moras. Kager (1995), too, notes that light monosyllables are very rare in Czech.

In my model data, I do not include monosyllabic words that start with a simple onset. However, the real language does contain monosyllables with simple onsets, which do not require a vocalized preposition. There must therefore be another constraint potentially interacting with MINWORD and *FEETUN, that punishes very complex onsets; this constraint is given in (30); it is based on the constraint *COMPLEX, which militates against associating more Cs or Vs than one to any syllable position (Prince & Smolensky 1993).

(30) Structural constraint

*ONSETCCC onsets are not composed of 3 or more segments

---

[29] The constraint *FeetUn could be rewritten as FeetBin ('feet are binary', Prince & Smolensky 1993) in which case a low-ranked constraint ParseSyll ('syllables must be parsed into feet', Prince & Smolensky 1993) would have to be employed as well to favor the stress pattern /σ́σσ/ (which is quite common in Czech) over /σ́σσ́/.

It is important to bear in mind that in the actual Czech grammar where syllables can be both light and heavy, and words can start with both simple and complex onsets, all the three constraints MINWORD, *FEETUN and *ONSETCCC must operate; all three of them are needed to explain the actual variation. In Tableaus (31) and (32) I suggest how the grammar can work and how the three newly introduced structural constraints interact.

Tableau (31) shows that the most appropriate evaluation strategy would be Harmonic Grammar (HG, Smolensky and Legendre 2006, Ch. 20). In HG, constraints are assigned weights. At an evaluation time, each candidate's weighted violations of the constraints sum up to give that candidate's harmony value. The most harmonic candidate, *i.e.* candidate with the highest harmony value, is selected as optimal. Tableau (31) consists of four sub-tableuas, in which I give pairs of possible candidates, *i.e.* always one candidate with vocalized and one with non-vocalized preposition. The constraint DEP is ranked highest and has a weight of 15, the weights of both MINWORD and *FEETUN are 10, and *ONSETCCC has a weight of 5 (it is ranked lowest of the four, because Czech allows onsets composed of up to 4 consonants).

(31) Tableau: The suggested HG evaluation strategy of 4 candidate-pairs: interaction of DEP, MINWORD, *FEETUN, and ONSETCCC.

constraint weight	*15*	*10*	*10*	*5*	
	DEP	MINWORD	*FEETUN	*ONSETCCC	*harmony*
☞ <to+us> |k+naːm| /.knaːm./			-1		-10
<to+us> |k+naːm| /.kɛ.naːm./	-1				-15
<to+dog> |k+psu| /.kpsu./		-1	-1	-1	-25
☞ <to+dog> |k+psu| /.kɛ.psu./	-1				-15
☞ <to+dogs> |k+psuːm| /.kpsuːm./			-1	-1	-15
☞<to+dogs>|k+psuːm|/.kɛ.psuːm./	-1				-15
☞ <to+dog>|k+psovɪ|/.kpso.vɪ./				-1	-5
<to+dog>|k+psovɪ|/.kɛ.pso.vɪ./	-1				-15

In Tableau (31) it is seen that the violation of *FEETUN alone cannot outrank the violation of DEP (therefore, monosyllables with simple onsets receive a non-vocalized preposition).

However, the violation of all the three lower-ranked constraints MinWord, *FeetUn and *OnsetCCC must outrank the violation of DEP (therefore a light monosyllable with a complex onset will always receive a vocalized preposition). It can be clearly seen (by replacing ‘-1’ with an asterisk) that if this tableau were evaluated with OT instead of HG, the grammar would fail to select the correct candidates and would always prefer the non-vocalized prepositions because of the high ranking of DEP.

In order to have OT be able to handle problems such as the one shown above, *local conjunction* (Smolensky 1997) must be put into operation. Under the local conjunction approach, two or more low ranked constraints can be conjoined to form a higher-ranked derived constraint, which is violated only if all the conjoined constraints are violated. If our three constraints MinWord, *FeetUn and *OnsetCCC formed a conjunction, *i.e.* a constraint WordFeetOnset, which would be ranked above DEP, then even an OT grammar would produce the correct output for light monosyllables as well as for simple-onset words. However, the present case is more complicated, because to have the grammar produce the correct output for heavy monosyllables, another conjunction would have to be made, namely that of *FeetUn and *OnsetCCC, *i.e.* FeetOnset, which would be ranked equally high as DEP. See Tableau (32) for an illustration of how this alternative analysis with local conjunctions works.

(32) Tableau: Evaluation with OT using local conjunction.

input to production	output of production	WordFeet Onset	FeetOnset	DEP	MinWord	*FeetUn	*OnsetCCC
	ranking value	*20*	*15*	*15*	*10*	*10*	*5*
<to+us>	☞ |k+na:m| /.kna:m./					*	
	|k+na:m| /.kɛ.na:m./			*!			
<to+dog>	|k+psu| /.kpsu./	*!			*	*	*
	☞ |k+psu| /.kɛ.psu./			*			
<to+dogs>	☞ |k+psu:m| /.kpsu:m./		*			*	*
	☞ |k+psu:m| /.kɛ.psu:m./			*			
<to+dog>	☞ |k+psovɪ| /.kpso.vɪ./						*
	|k+psovɪ| /.kɛ.pso.vɪ./			*!			

In this study I consider all monosyllabic words together and ignore the difference between heavy and light monosyllables, as well as the difference between monosyllables with simple and complex onsets. The model that I create in my study is not an attempt at whole-language simulations, and is primarily aimed at showing how a listener-oriented grammar explains vocalizations in the prepositions. Hence, other effects such as those resulting from interactions between the three structural constraints MINWORD, *FEETUN and *ONSETCCC are not included in the model. Nevertheless, I have tried to suggest how these other constraints and their rankings could predict the observed language data. Importantly, I show that a constraint like MINWORD affects prepositional vocalizations. To the best of my knowledge, such an observation has not been made in previous studies.

Because, from now on, I ignore heavy monosyllables and monosyllables with simple onsets, I will use only the constraint MINWORD in my model; see Tableau (33) that illustrates how the grammar with this high-ranked constraint works.

(33) Tableau: Production of preposition + a light monosyllable with a complex onset

<to+dog>	*[x] / /	*[ ] /x/	MINWORD	*[7cue] /CCCC/	DEP	*[6cue] /CCC/	*[7cue] /CCC/
|k+psu| /.kpsu./ [kpsu]			*!			*	
|k+psu| /.kpsu./ [kεpsu]	*!		*				
☞ |k+psu| /.kε.psu./ [kεpsu]					*		
|k+psu| /.kε.psu./ [kpsu]		*!					

The last two word-types that remain to be addressed in the present analysis are words represented here by *auto* and *rtuti*. The former always occurs with non-vocalized prepositions, while the latter usually occurs with vocalized prepositions.

4.10 Vowel-initial words

Recall that vowel-initial words surface with a glottal stop preceding the vowel; this happens especially if they are preceded by a pause or by a non-syllabic preposition (Hála 1962:280). Palková (1994:325-6) notes that the glottal stop is usually used when a vowel-initial word or word stem is preceded by an unstressed monosyllabic word, or when the preceding word or prefix ends in a vowel, while the absence of a glottal stop before a vowel-initial word in any environment seems to be a result of a less careful, fast speaking style. Palková (1994:325) also states that if a glottal stop does not occur, the final consonant of the preceding word is never resyllabified into the onset of the vowel-initial word. However, it remains to be investigated experimentally whether there is any evidence in the phonetic form of such words that no resyllabification of the final consonant into the onset takes place. If it then turns out that there is indeed no evidence for resyllabification, it must be examined whether another segment or a phonetic event (such as a voiceless /h/ or a short creak) marks the onset of the vowel-initial word instead of a fully realized glottal stop.

In Section 4.8 I have suggested that the grammar may prefer a complete syllabification of the non-vocalized prepositions into the following onset, similarly to Rubach (2000), and contrary to the pre-syllabic status of these prepositions proposed by Kučera (1961); see Section 3.3 of the present study. If the preposition is fully syllabified into the onset, glottal stop insertion (which is obligatory with a non-vocalized preposition) cannot be seen as satisfying the requirement for an onset in Czech; this is because the preposition could function as an onset and no extra segment such as a glottal stop would have to be inserted to satisfy the onset requirement. Rubach (2000) puts forth a theory of Derivational OT and proposes that the glottal stop is inserted at the stage of word morphology; word morphology is optimized earlier than phrase morphology and therefore the prepositions are pre-posed to a word that already starts with a glottal stop.

Contrary to previous analyses (such as Rubach 2000) that assumed glottal stop insertion during the production of 'vowel-initial' words, in the present study it is argued that the glottal stop is represented underlyingly in the words that orthographically start in a vowel; in fast sloppy speech the glottal stop may be deleted during phonetic implementation if at least one consonant precedes, because of articulatory constraints that disfavor clusters with a glottal stop; this articulatory constraint was proposed by Boersma (2007) and is described in (34). The glottal stop will never be deleted if the deletion would result in miscomprehending the message, as would often be the case with non-syllabic prepositions, as shown in the Tableaus (35a) through (35c). In these Tableaus, it is clearly seen that in the adult grammar, lexical constraints are either low-ranked (those that ban the correct UF) or high-ranked (those that ban an incorrect UF).

(34) Articulatory constraint

*[CʔC]$_{\mathrm{Art}}$ do not realize a glottal stop in the context of (an)other consonant(s)

(35a) Tableau: Faulty comprehension due to glottal stop deletion in a prep. phrase

<table>
<tr><th>[kuːlu]</th><th>*/ / [x]</th><th>*/x/ [ ]</th><th>*[CʔC]$_{\mathrm{art}}$</th><th>*<to> |k|</th><th>*<hive> |ʔuːlu|</th><th>*<stake> |kuːlu|</th></tr>
<tr><td>/.k ʔuː.lu./ |k+ʔuːlu| <to+hive></td><td></td><td>*!</td><td></td><td>*</td><td>*</td><td></td></tr>
<tr><td>☞ /.k uː.lu./ |kuːlu| <stake></td><td></td><td></td><td></td><td></td><td></td><td>*</td></tr>
</table>

(35b) Tableau: Correct comprehension in a prepositional phrase

<table>
<tr><th>[kʔuːlu]</th><th>*/ / [x]</th><th>*/x/ [ ]</th><th>*[CʔC]$_{\mathrm{art}}$</th><th>*<to> |k|</th><th>*<hive> |ʔuːlu|</th><th>*<stake> |kuːlu|</th></tr>
<tr><td>☞ /.k ʔuː.lu./ |k+ʔuːlu| <to+hive></td><td></td><td></td><td>*</td><td>*</td><td>*</td><td></td></tr>
<tr><td>/.k uː.lu./ |kuːlu| <stake></td><td>*!</td><td></td><td></td><td></td><td></td><td>*</td></tr>
</table>

(35c) Tableau: Correct comprehension despite glottal stop deletion

<table>
<tr><th>[daloko]</th><th>*<tree> |oko|</th><th>*[x]/ /</th><th>*[ ]/x/</th><th>*[CʔC]$_{\mathrm{art}}$</th><th>*<gave> |dal|</th><th>*<eye> |ʔoko|</th></tr>
<tr><td>☞/.dal.ʔo.ko./ |dal+ʔoko| <gave+eye></td><td></td><td></td><td>*</td><td></td><td></td><td>*</td></tr>
<tr><td>/.dal.o.ko. / |dal+oko| <gave+tree></td><td>*!</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

In the Tableaus above, lexical constraints that are low-ranked are those that militate against the correct UF, *i.e.* in Czech *úlu* means ‘hive+Gen.’ *dal* means ‘(he)gave’, *oko* means ‘an eye’, hence the low-ranked constraints *<hive> |ʔuːlu|, *<gave> |dal|, *<eye> |ʔoko|. Analogously, the high-ranked lexical constraints militate against nonsense morpheme-UF mappings, *i.e.* if *oko* is pronounced without the glottal stop it does not mean e.g. ‘tree’ in Czech, thus a high-ranked constraint such as *<tree> |oko|.

In this section I have shown a plausible analysis with underlying glottal stops. It will be seen in the results of the learning simulations (Section 5) whether learners choose the glottal stop to be present in the UF or whether they insert it during production.

4.11 Sonorant-obstruent onsets

The case of *rtuti* is in the present study analyzed by means of McCarthy’s (1998) Sympathy. Let us first attempt an analysis without Sympathy.

The OT grammar described thus far would mostly prefer the candidate |k+rtucɪ| /.krtu.cɪ./ [krtucɪ] but this is not what we see in most of the language data. In reality, the preferred form is /.kɛ.rtu.cɪ./, and /.krtu.cɪ./ is only found occasionally in the speakers’ output. It has been reported for Czech that a sonorant which is word-final and preceded by at least one consonant, or is between two consonants word internally, is always syllabic (Palková 1994:270). This is formalized with the structural constraint that I give in (36).

(36) Structural constraint

*/OSO/	a sonorant between two obstruents word-internally is not non-syllabic

An analysis with this new high-ranked constraint would prefer syllabic /r/ in the winning form: /.kŕ̩.tu.cɪ./. Since in reality if the preposition is non-vocalized, stress is placed on /u/ and not on /r/, the SF /.kŕ̩.tu.cɪ./ should not win, as is the case in Tableau (37).

(37) Tableau: production of |k+rtucɪ| without Sympathy.

|k+rtucɪ|	*/OSO/	*[7cue] /CCCC/	DEP	*[6cue] /CCC/	*[7cue] /CCC/	ALIGN
☹ /.krtú.cɪ./ [krtʊcɪ]	*!			*		**
☞ /.kŕ̩.tu.cɪ./ [kɾtucɪ]						**
☹ /.kɛ́.rtu.cɪ./ [kɛrtucɪ]			*!			*

Tableaus (38) and (40) illustrate a grammar that employs Sympathy. First, a sympathetic candidate is selected; this is done by intuitively and informally considering the advantage of faithfulness that the correct candidate has over the others to this sympathetic candidate (Kager 1999:388). The sympathy in the present case lies in preserving the consonantal non-syllabic status of the word-initial liquid. Therefore the sympathetic candidate will be the one that is faithful to the UF and violates the constraint */OSO/, *i.e.* /.krtú.cɪ./ [krtʊcɪ] denoted by the symbol ❀. Tableau (38) shows that the constraint punishing unfaithfulness to the sympathetic candidate – this constraint is described in (39) – has to be ranked above DEP, to have the grammar select the candidate with SF /.kɛ́.rtu.cɪ./ as optimal.

(38) Tableau: production of |k+rtucɪ| with Sympathy.

|k+rtucɪ|	*/OSO/	*[7cue] /CCCC/	*❀/C/ =/V/	DEP	*[6cue] /CCC/	*[7cue] /CCC/	ALIGN
❀ /.krtú.cɪ./ [krtʊcɪ]	*!				*		**
/.kŕ̩.tu.cɪ./ [kɾtucɪ]			*!				**
☞ /.kɛ́.rtu.cɪ./ [kɛrtucɪ]				*			*

(39) Sympathy constraint

*❀/C/ = /V/ — a segment which is a consonant in the SF of the sympathetic candidate does not map to a vowel in the SF of the optimal candidate

The problem is more complex, however, because some variation is observed with the *rtuti*-type words. The prepositions are sometimes non-vocalized. In other words, sometimes the winner is the sympathetic candidate itself. We can model this with Stochastic OT. The constraint *❁/C/ = /V/ will remain ranked above DEP, but will now outrank */OSO/ as well. The ranking value of */OSO/ will be slightly higher than the ranking value of DEP so that at some evaluation times, the candidate with the SF /.krtú.cɪ./ will win; this is formalized in Tableau (40).

(40) Tableau: variability in vocalizations in the production of |k+rtucɪ|.

ranking value	100	82	80.5	80	80.1	78	50	frequency of this winner
|k+rtucɪ|	*❁/C/ = /V/	*[7cue] /CCCC/	*/OSO/	DEP	*[6cue] /CCC/	*[7cue] /CCC/	ALIGN	
❁ /.krtú.cɪ./ [krtucɪ]			*!		*		**	25%
/.kŕ̩.tu.cɪ./ [krtucɪ]	*!						**	
☞ /.kɛ́.rtu.cɪ./ [kɛrtucɪ]				*			*	75%

Tableau (40) shows that the grammar now produces the SO-initial words with the vocalized form of the preposition [kɛrtucɪ] in most cases; these sometimes alternate with the non-vocalized forms, in which the sonorant stays non-syllabic [krtucɪ]. This agrees with what has been described in Section 2.2.

Note that instead of using Sympathy, the problem could as well be solved by *output-to-output correspondence* (‘OO-correspondence’, e.g. McCarthy 1995, Kenstowicz 1996). The OO-correspondence is based on the notion that some constraints operate between morphologically related forms and such constraints require faithfulness between the related forms (‘paradigm uniformity’). In the present analysis then, on the basis of the existence of forms such as /.rtu.cɪ./ ‘mercury-DAT’, /.rtu.ciː./ ‘mercury-INSTR’, or /.rtu.co.viː./ ‘mercury (Adj.)’, whose base /.rtuc./ contains a non-syllabic /r/, the OO-faithfulness constraints would require that, also in cases in which these words are preceded by a non-syllabic preposition, the /r/ remains non-syllabic, *i.e.* /.krtu.cɪ./ or /.kɛ.rtu.cɪ./.

It is seen that in both of these analyses, the main point is preserving faithfulness (more specifically consonantal identity) in the onset-initial sonorant /r/. It is not the aim of the present study to investigate which of the two analyses is most suitable for the vocalizations in sonorant-obstruent clusters; I therefore pick one of the two, which is Sympathy.

4.12 The final grammar

This section puts together all the 24 constraints that were employed in 4.3 through 4.11 to explain the grammar of Czech prepositional vocalizations. The final formalization of the whole grammar, which was modeled in steps throughout Section 4, is presented in (41). The constraint labels are listed with their *relative* rankings, which, I argue, are crucial to produce forms that speakers of Czech produce.[30]

(41) Constraint list

(a)	Lexical constraints	ranking value
	*<to> |k|	20
	*<to> |kɛ|	120
(b)	Faithfulness constraints	
	MAX	100
	DEP	80
	IDword	100
	IDprep	50

[30] For the description of the constraints, I refer to the section (4.3 through 4.11) in which the respective constraint was introduced.

(c) Structural constraints

VOI	100
PARSE	100
*O/N	100
ALIGN	50
MINWORD	100
STRESS 1st	100
*/OSO/	81.5

(d) Cue constraints

*/-,+/ [+,-]	100
*/CC/ [_C:]	100
*/ / [x]	100
*/x/ [ ]	100
*/a , á/ [ɑ , a]	100
*/CCCC/ [7cue]	82
*/CCC/ [6cue]	80.1
*/CCC/ [7cue]	78

(e) Articulatory constraints

$*[C_iC_i]_{Art}$	100
$*[C\text{ʔ}C]_{Art}$	95

(f) Sympathy constraint

*❀/C/ = /V/	100

5 Learning simulations

5.1 Aims of the simulations

In the previous section, I showed that the grammar created with the BiPhon model and Stochastic OT can successfully explain the prepositional vocalizations (with some refinements such as local conjunction or HG, and Sympathy or OO-correspondence). Importantly, the model accounts for the observed variation as well. With the constraint rankings that I set at certain values, virtual speakers produce such forms that are attested in the reality. The model is a formalization of an adult grammar, *i.e.* of speakers that had already acquired the grammar, *i.e.* language-specific constraint rankings. It remains to be seen whether the linguistic analysis of Section 4 is learnable, which is the aim of the present section.

At the initial state, before learning commences, all constraints have equal ranking values. The virtual children learn from overt forms, *i.e.* they learn from the auditory forms that Czech speakers produce. If, after having received a sufficient amount of input, the learners re-rank the constraints in such a way that the prepositional forms they themselves produce and the frequency of the vocalizations with various onsets are adult-like, *i.e.* if their output is comparable to the adult overt forms that they learned from, learning will have been successful and the language (*i.e.* prepositional vocalizations of Czech) will have been proven to be learnable.

Next, the simulations will answer several crucial questions, which, up to now, have merely been an object of speculation. One goal is to find out what the underlying form of the prepositions is; outcomes of the learning simulations will show whether learners choose |k| as the underlying form, as I assumed in Section 4, or |kɛ|, or both (*i.e.* if they have two allomorphs underlyingly). The next issue that will be answered is the pattern of syllabification of the non-vocalized prepositions. The learners will be free to choose between not parsing the consonant into a syllable at all (and thus analyzing it as a 'pre-syllabic' segment), parsing the consonant in such a way that it forms a syllable by itself, and fully parsing the consonant into the onset of the following word. If the learners will consistently tend to choose one of these three syllabification strategies, then that strategy is likely to be the same one that real human speakers employ. Last, it will be seen whether 'vowel-initial' words start in a vowel underlyingly and the glottal stop is inserted during production, or whether the glottal stop is already present in the underlying form.

5.2 The learning strategy

The learning simulations are run with the computer program Praat (Boersma & Weenink 2009). Learners learn from overt forms, *i.e.* the only information that is available to them during learning is the distribution of the real language data. From this distribution a total of 400 000 overt forms are drawn equally divided up between four different plasticity stages; the plasticity starts at 1 and decreases by a factor of 0.1 at each step. From the overt forms, the learners have to reconstruct the most optimal triplets SF-UF-morpheme. In fact, for a given overt form the learners have to compute only the SF and the UF; it is rather trivial for them to arrive at the morpheme, because each particular overt form in their input corresponds to one morpheme. In other words, once the learning algorithm finds the tableau that contains a given overt form, it has found the morpheme for this overt form as well.

After the learners map an overt form to the morpheme, they compute which triplet of UF-SF-OF they would themselves produce for this morpheme (‘virtual production’, described in detail in Apoussidou & Boersma 2004). They compare this virtually derived form to the adult form and if the two forms differ, they re-rank the respective constraints. The strategy they employ is a version of the Gradual Learning Algorithm (Boersma 1998, Boersma & Hayes 2001); they raise the ranking of all constraints that are violated in the form they would produce and lower the ranking of all constraints that are violated in the adult form; the size of such a learning step is inversely proportional to the number of re-ranked constraints.[31]

5.3 Learning prepositional vocalizations

Ten learners learned from 10 tableaus, one for each of the 10 morphemes that represent the real language data as listed in Table (6), Section 4.2. The distribution of adult overt forms,

[31] In Praat this strategy is called ‘weighted all’.

from which the learners learn (42) is based on a corpus analysis of word types occurring with the vocalized and non-vocalized prepositional forms.[32]

(42) THE DISTRIBUTION OF ADULT OVER FORMS: THE OVERT FORM + ITS FREQUENCY

"[ktiːmu]"	165	"[gdomu]"	165	"[kʔautu]"	47
"[kautu]"	3	"[kɛkolu]"	10	"[kɛgoːlu]"	10
"[kɛpsu]"	8	"[kplotu]"	27	"[kpsovi]"	15
"[kʃkviːr̝ɛ]"	3	"[krtuti]"	2	"[kɛpsovi]"	15
"[kɛʃkviːr̝ɛ]"	27	"[kɛplotu]"	3	"[kɛrtuti]"	6

The 10 tableaus contained 247 candidates in total (each candidate was a triplet UF-SF-OF), including 'sanity candidates' so that the learners were able to figure out that some constraints are likely to be ranked high in the language (e.g. the constraint MAX must be ranked quite high because it is the only constraint that a candidate such as "|k+tiːmu| /.kiː.mu./ [kiːmu]" violates but we never find mappings <to+team> → [kiːmu] in the adult output).

The number of constraints that were used was 26. Besides the 24 constraints listed in Section 4.11, two more lexical constraints were necessary to account for learning of the

[32] In the SYN2005 corpus, there are in total 509 000 instances of the preposition <to> (all numbers I report here are rounded thousands), 80 000 of which are *ke* and 429 000 are *k*. The form *ke* is in 20 000 cases followed by a word that starts in an identical consonant, and in 60 000 cases it is followed by a complex onset. The form *k* is in 330 000 cases followed by a word starting in a simple onset and in 50 000 cases by a 'vowel-initial' word; in 49 000 cases it is followed by a complex onset. The numbers shown in the distribution (42) were derived from the corpus values as well as from the generalizations on variation that I made in Section 2 (*i.e.* that approximately 90% of very complex onsets are preceded by *ke*, 90% of the least complex onsets are preceded by *k*, while the intermediate complex onsets are preceded by both prepositions equally often; and that light monosyllables with complex onset are always preceded by *ke*). More detailed corpus analyses of all the four prepositions and of all the possible onsets and of all word sizes will be necessary to replicate faithfully the situation that takes place in the real language.

‘vowel-initial’ words (43); if the learners choose to rank *<car> |auto| higher than *<car> |ʔauto|, the UF of ‘vowel-initial’ words will contain the glottal stop (potentially also in real language users), if on the other hand they rank the two constraint in the reverse order, the glottal stop will not be present underlyingly (potentially also in real language users).

(43) Lexical constraints

*<car> |auto|	do not connect the morpheme <car> with the UF |auto|
*<car> |ʔauto|	do not connect the morpheme <car> with the UF |ʔauto|

Appendix 1 shows the complete grammar before learning with all constraints ranked equally high; Appendix 2 contains the distribution of adult overt forms, which together with the initial grammar were used to run the learning simulations; the script for learning simulations is shown in Appendix 3.

5.4 Results

The prepositional vocalizations were learned successfully by all 10 learners. Table (44) shows what constraint rankings each of the 10 learners arrived at. Since all learners converged towards the same language, an average ranking value is given for each constraint. If the final OT grammar is evaluated 1000 times with a random noise drawn from a Gaussian distribution with a standard deviation of 2 around the ranking values, it is seen what the virtual learners’ output is (45).

(44) Table: The acquired grammar.

constraint	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	ranking value[33]
*<to> |ke|	109.1	109.7	110.2	105.7	110.4	109.4	111.5	110.5	109.0	109.5	**109.5**
*/ / [x]	107.8	108.3	109.0	108.1	108.2	108.2	108.4	109.2	108.2	107.3	**108.3**
*<car> |auto|	107.7	108.7	106.9	107.2	106.9	108.2	107.1	108.4	106.9	107.3	**107.5**
*[CiCi]$_{\text{Art}}$	107.0	107.0	108.7	106.7	108.0	107.9	108.3	107.7	107.1	106.6	**107.5**
*/CC/ [_C:]	107.0	106.7	108.6	106.4	107.6	107.3	107.5	107.8	107.0	106.2	**107.2**
MINWORD	105.9	106.4	107.5	105.3	106.8	105.9	106.5	106.7	106.2	105.3	**106.2**
* ❁ /O/=/S/	104.8	106.2	106.7	105.8	106.3	105.9	106.2	106.2	105.9	105.5	**105.9**
*O/N	106.0	105.0	107.8	103.9	105.7	104.5	106.7	106.0	105.4	103.7	**105.5**
VOI	101.9	103.7	104.2	102.9	103.2	103.0	103.6	102.8	102.1	101.9	**102.9**
*/CCCC/ [7cue]	100.9	102.3	102.8	101.9	102.1	102.3	102.0	102.2	102.1	101.3	**102.0**
PARSE	101.1	101.8	101.9	101.8	101.8	100.4	101.0	100.0	103.0	101.2	**101.4**
IDword	101.3	100.5	100.2	100.2	99.4	100.8	100.8	100.8	100.1	101.1	**100.5**
*/OSO/	99.4	100.4	101.0	100.5	100.4	101.2	100.4	100.2	100.5	99.9	**100.4**
/á, a/ [a, **a**]	99.6	98.9	97.3	99.1	100.0	100.0	99.1	99.6	99.8	99.4	**99.3**
*/x/ [ ]	99.1	96.4	99.2	98.9	100.7	97.6	100.8	98.1	99.2	98.9	**98.9**
*<car> |ʔauto|	99.1	99.2	98.3	98.2	98.3	98.7	97.9	98.7	97.9	98.4	**98.5**
*/CCC/ [6cue]	97.3	98.7	99.2	98.3	98.4	98.9	98.5	98.6	98.6	97.8	**98.4**
DEP	97.3	98.6	99.2	98.3	98.4	98.8	98.4	98.6	98.5	97.8	**98.4**
MAX	99.2	96.9	99.0	99.1	100.2	98.1	93.0	98.2	99.3	98.9	**98.2**
*<to> |k|	96.9	98.6	99.2	96.4	96.6	96.9	99.1	98.7	96.9	96.1	**97.5**
*/+,-/ [-,+]	98.8	98.3	89.9	100.7	86.6	99.1	100.9	97.9	93.8	99.5	**96.6**
*/CCC/ [7cue]	93.6	94.9	95.5	94.7	94.8	95.1	94.8	95.0	94.8	94.1	**94.7**
STRESS 1st	96.2	90.9	84.6	93.6	93.7	92.4	101.2	92.8	91.7	96.2	**93.3**
*[C?C]$_{\text{Art}}$	93.9	91.6	94.0	93.9	95.2	92.6	88.5	93.0	94.0	93.5	**93.0**
ALIGN	86.9	85.2	85.7	87.4	87.9	86.8	80.5	87.0	85.3	86.8	**85.9**
IDprep	82.2	85.0	83.5	85.2	82.6	80.0	77.5	75.3	86.7	85.7	**82.4**

[33] Constraints with ranking values far apart yield fixed dominance (ranking difference of approx. 10), constraints with nearby ranking values yield variation (Boersma & Hayes 2001).

Table (45) shows that the language was learnt correctly because the learners' output resembles the adult output they learnt from. Table (44) shows that the analysis proposed in Section 4 (*i.e.* the crucial relative constraint rankings) is what the learners' final grammar resembles to.

That is, for example the constraints *[CʔC]art and IDprep are ranked low; the grammar thus favors glottal stop deletion in a consonant cluster (e.g. in fast speech), and voicing assimilation of the preposition to the following-word onset, respectively. Importantly, the cue constraints I introduced that handle the vocalizations and the variation are ranked as I had proposed; */CCCC/ [7cue] is ranked high so that very complex and similar onsets almost always take the vocalized preposition; this constraint also outranks */CCC/ [6cue], which in turn outranks */CCC/ [7cue]. The constraint */CCC/ [6cue] is ranked equally with DEP allowing for the 50:50 variation in vocalizations with moderately complex onsets. The constraint */CCC/ [7cue] is ranked below DEP, so that the least complex onsets favor the non-vocalized forms of the prepositions. The constraints $*[C_iC_i]_{Art}$ and */CC/ [_C:] are ranked very high, which means that the identity between the preposition and the consonant will favor /ɛ/ insertion (thanks to a low ranking of DEP).

Because the language was shown to be learnable and because all 10 learners converged to the same language I will now use the final grammar to find answers to the questions that I asked about the 'hidden structures', *i.e.* the UF and the SF. This is done in the following section.

(45) Produced forms after learning[34]

morpheme	UF-SF-OF output	frequency of this output (%)
<to+a team>	|k+tiːmu| /.ktíː.mu./ [ktiːmu]	100
<to+a house>	|k+domu| /.gdó.mu./ [gdomu]	100
<to+a bike>	|k+kolu| /.kɛ́.ko.lu./ [kɛkolu]	100
<to+a goal>	|k+goːlu| /.kɛ́.goː.lu./ [kɛgoːlu]	99.9

[34] Learner output forms with frequencies less than 1% are not included in the table.

<to+a car>	|k+ʔautu| /.kʔáu.tu./ [kʔ**ɑʊ**tu] |k+ʔautu| /.kʔáu.tu./ [k**ɑʊ**tu] |k+ʔautu| /.káu.tu./ [k**ɑʊ**tu]	95.3 1.6 2.9
<to+a chink>	|k+ʃkviːr̝ɛ| /.kɛ́.ʃkviː.r̝ɛ./ [k**ɛ**ʃkviːr̝ɛ] |k+ʃkviːr̝ɛ| /.kʃkvíː.r̝ɛ./ [kʃkv**iː**r̝ɛ]	90.2 9.8
<to+a dog1>	|k+psu| /.kɛ́.psu./ [k**ɛ**psu]	99.2
<to+a dog2>	|k+psovɪ| /.kɛ́.pso.vɪ./ [k**ɛ**psovɪ] |k+psovɪ| /.kpsó.vɪ./ [kps**o**vɪ]	51.5 48.5
<to+a fence>	|k+plotu| /.kɛ́.plo.tu./ [k**ɛ**plotu] |k+plotu| /.kpló.tu./ [kpl**o**tu]	9.3 90.7
<to+mercury>	|k+rtucɪ| /.kɛ́.rtucɪ./ [k**ɛ**rtucɪ] |k+rtucɪ| /.krtú.cɪ./ [krt**u**cɪ]	73.4 26.1

6 Conclusions

6.1 Hidden structures

In Section 4, I proposed an analysis that should explain prepositional vocalizations in Czech. Section 5 then tested whether this analysis is plausible. Apart from showing whether the proposed analysis is learnable, Section 5 answered three major questions regarding the underlying and surface structures.

First, I aimed at finding out what the UFs of the prepositions are. As Table (44) shows, the constraint *<to> |kɛ| is ranked high. This implies that the UF of the prepositions in my model is the non-vocalized form (as can be also seen in the learners' output in (45)). The standpoint that the majority of the literature and previous studies seem to hold, as well as the assumption that I made in Section 4, which was based on corpus frequencies of the two prepositional allomorphs, that the underlying form is non-vocalized, is thus confirmed.

Second, the learners were given the opportunity to choose not to parse the consonantal segment into a syllable at all (keep it a 'pre-syllabic' element), or to syllabify it as a single whole syllable, or to make it a part of the following onset. The low ultimate ranking of ALIGN suggests that the learners arrive at a preference for the last option of the three. Indeed, the syllabification pattern that the learners produced was to align the non-vocalized consonantal preposition to the onset of the following word. I therefore conclude that it is not necessary to invent constructs such as ‘pre-syllables’ when referring to non-vocalized prepositions.

The last question that was addressed was whether the learners choose ‘vowel-initial’ words to start in a glottal stop underlyingly, or whether they insert the glottal stop during production. All of them ranked the constraint *<car> |auto| high. The learners’ underlying form of ‘vowel-initial’ words contains a glottal stop word-initially. As Table (45) also shows, the glottal stop is in the UF and is occasionally deleted during production (probably because of fast speaking style).[35] Since the learners in this study ranked *<car> |auto| very high, I conclude that it is highly plausible that human learners do so, too, and have ‘vowel-initial’ words represented with a glottal stop underlyingly.

6.2 Always-syllabic prepositions

In Section 2 I have proposed, contrary to previous assumptions that vocalizations in non-syllabic and always-syllabic prepositions are handled by different principles, that the present model created on the basis of the distribution of non-syllabic prepositions will account for vocalizations in the other prepositions as well. In Tableaus (46) through (48), it is seen that my model does indeed explain the (non-)vocalizations in other prepositions, too. Crucially, these other prepositions are always syllabic and thus their final consonant is always a coda. Because the final consonant of the always-syllabic preposition is always preceded by a vowel, the constraint */CC/ [_C:] is never violated. The fatal violation of DEP punishes the vocalized candidates and the non-vocalized preposition is selected as optimal, even if the following word begins in a consonant identical to the final consonant of the preposition.

[35] Note that for the present learners the deletion could happen both in the SF as well as in the OF. This is because MAX and */x/ [ ] received similar ranking values. It would be necessary to present the learners with more language data to find out where the glottal stop deletion happens. The constraint */x/ [ ] could be split into */V/ [ ] and */ʔ/ [ ], and eventually the latter would be ranked lower than the former and also lower than MAX. This is what I assume on the basis of the fact that a glottal stop is less prominent auditorily than e.g. a vowel. In such a case, the glottal stop deletion would happen only in the OF.

Light monosyllables do not require vocalization of the preposition because the resulting prosodic item is bisyllabic and does not violate the high-ranked MINWORD. Since the final consonant is a coda of the prepositional syllable, the constraint */OSO/ that applies *within* but not across words is not violated either if the pre-modified word has an onset of the structure /SO/.

A search in the corpus confirms that the forms predicted by the grammar are indeed attested in the language. The always-syllabic prepositions are almost never vocalized except for idiomatized expressions, such as in *ode dne ke dni*[36] ‘day by day’, *beze strachu* ‘with no fear’, or in the oblique cases of the pronoun *já* ‘I’ (*i.e. mne, mě, mně, mi, mnou*).

(46) Tableau: Non-vocalization of a syllabic preposition when an identical onset follows.

<table>
<tr><td>ranking value</td><td>109.5</td><td>107.2</td><td>98.4</td><td>98.2</td><td>97.5</td><td>85.9</td></tr>
<tr><td><without+winter></td><td>*<without>
|bɛzɛ|</td><td>*/CC/
[_C:]</td><td>DEP</td><td>MAX</td><td>*<without>
|bɛz|</td><td>ALIGN</td></tr>
<tr><td>☞|bɛz+zɪmɪ| /.bɛz.zɪ.mɪ./
[bɛzːɪmɪ]</td><td></td><td></td><td></td><td></td><td>*</td><td></td></tr>
<tr><td>|bɛz+zɪmɪ| /.bɛ.zɛ.zɪ.mɪ./
[bɛzɛzɪmɪ]</td><td></td><td></td><td>*!</td><td></td><td>*</td><td>*</td></tr>
<tr><td>|bɛzɛ+zɪmɪ| /.bɛ.zɛ.zɪ.mɪ./
[bɛzːɪmɪ]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>|bɛzɛ+zɪmɪ| /.bɛz.zɪ.mɪ./
[bɛzzɪmɪ]</td><td>*!</td><td></td><td></td><td>*</td><td></td><td>*</td></tr>
</table>

[36] However, there are a lot of instances of *ode dne* ‘from a day’, and not so many cases of *od dne*. It seems that the two identical plosives followed by a homorganic nasal is a very strong constraint (*i.e*. ‘too few different auditory cues do not map onto three segments in the SF’) against perceiving three different segments and force the /ɛ/-insertion.

(47) Tableau: Non-vocalization of a syllabic preposition when a complex onset follows.

ranking value	109.5	106.2	98.4	98.2	97.5	85.9
<without+dog>	*<without> |bɛzɛ|	MIN WORD	DEP	MAX	*<without> |bɛz|	ALIGN
☞ |bɛz+psa| /.bɛs.psa./ [bɛspsa]					*	
|bɛz+psa| /.bɛ.zɛ.psa./ [bɛzɛpsa]			*!		*	*
|bɛzɛ+psa| /.bɛ.zɛ.psa./ [bɛzpsa]	*!					
|bɛzɛ+psa| /.bɛs.psa./ [bɛspsa]	*!			*		*

(48) Tableau: Non-vocalization of a syllabic preposition when a /SO/ onset follows.

ranking value	109.5	100.4	98.4	98.2	97.5	85.9
<without+mercury>	*<without> |bɛzɛ|	*/OSO/	DEP	MAX	*<without> |bɛz|	ALIGN
☞ |bɛz+rtucɪ|/.bɛz.rtu.cɪ./ [bɛzrtucɪ]					*	
|bɛz+rtucɪ|/.bɛ.zɛ.rtu.cɪ./ [bɛzɛrtucɪ]			*!		*	*
|bɛzɛ+rtucɪ|/.bɛ.zɛ.rtu.cɪ./ [bɛzrtucɪ]	*!					
|bɛzɛ+rtucɪ|/.bɛz.rtu.cɪ./ [bɛzrtucɪ]	*!			*		*

6.3 Summary of the findings

This study has explained vocalizations of non-syllabic prepositions in Czech. It has provided a full account of what determines whether non-syllabic prepositions appear in their vocalized or in their non-vocalized form. Moreover, this study also accounts for the widely attested variation in the vocalizations.

The main idea behind the model I have proposed is that when a vocalized preposition occurs, it is for the listener to be able to recover the preposition, and not for the speaker to facilitate articulation, as had been assumed previously. If the number of distinctive auditory cues in the onset is low, the listener might not always be able to comprehend the message well enough. If then the number of distinctive auditory cues is too low, vocalization emerges. If the vocalization is present, the preposition will never be ‘lost’.

I also showed that, in words with complex onsets, it is not only the following onset structure that affects the vocalizations but also the length of the whole resulting sequence formed by the preposition and the pre-modified word.

Conclusions were made about how the prepositions are represented underlyingly, and about the most likely syllabification pattern with respect to the degree of alignment of the preposition to the following onset. The results of the learning simulations have also suggested that for Czech speakers, a glottal stop is present underlyingly as an initial segment of words that in their spelled forms start in a vowel.

The model that I proposed was originally based on the behavior of non-syllabic prepositions but it was shown that this model can explain vocalizations in the always-syllabic prepositions as well, for which – as I showed – no separate mechanism is needed.

6.4 Future research

Looking at the final grammar, it is apparent that something has to be done with the Sympathy constraint *❀/C/ = /V/ (introduced in Section 4.10). Even though the language was learnable, this constraint was asserted somewhat artificially to be able to handle the special cases of words in which neither onset complexity, onset similarity, nor the length of the word can explain the vocalizations. It might be the case that there are two different phonemes of each of the sonorants underlyingly, and that a candidate with a SF such as /.kŕ̩.tu.cɪ./, instead of violating the special constraint *❀/C/ = /V/, simply violates a faithfulness constraint because the /r/ is underlyingly specified as a consonant |rtucɪ| = |CCVCV| and would then have to change to a vowel in /.kŕ̩.tu.cɪ./ = /.CV.CV.CV./. I suggested that a possible alternative to the analysis with Sympathy could be OO-correspondence. It remains to be found which of the OT analyses that deal with phonological opacity is most plausible for the problem addressed in this study.

The present study has claimed, based on the performance of virtual speakers, that non-syllabic prepositions are fully aligned to the following onset, if non-vocalized. Results of experiments with human subjects could show whether such a conclusion is plausible or not. If, for example, the onset of a word such as *Květě* 'female name, Dat.' /kvjɛcɛ/ is the same as the onset formed by a preposition and a word such as *k větě* 'to a sentence' /kvjɛcɛ/, the claim made in the present study would be strongly supported. The similarity should then be assessed both by perception experiments as well as by analyses of productions of real language users.

This study has shown that many issues related to prepositional vocalizations can be explained by a simple grammar that contains 25 cross-linguistically attested constraints (*i.e.* except Sympathy), which – if ranked in the way in which the virtual learners ranked them based on the distributions of the real language data – can not only predict in what environments the vocalizations will occur but can also account for the variation that is present in the productions of Czech speakers.

References

Apoussidou, Diana (2007). *The Learnability of Metrical Phonology*. PhD dissertation, University of Amsterdam.

Apoussidou, Diana and Paul Boersma (2004). Comparing Two Optimality-Theoretic Learning Algorithms for Latin Stress. In Schmeiser, B., V. Chand, A. Kelleher and A. Rodriguez (eds), *WCCFL 23 Proceedings*. Somerville, MA: Cascadilla Press, 101-114.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *IFA Proceedings* 21. 43–58.

Boersma, Paul (1998). *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual drives.* PhD dissertation, University of Amsterdam.

Boersma, Paul (2000). The OCP in the perception grammar. [*Rutgers Optimality Archive:* ROA-435]

Boersma, Paul (2007). Some listener-oriented accounts of h-aspiré in French. *Lingua* 117: 1989-2054.

Boersma, Paul (2009). Cue constraints and their interactions in phonological perception and production. To appear in Paul Boersma & Silke Hamann (eds.). *Phonology in perception*. Berlin: Mouton de Gruyter. [ROA-944]

Boersma, Paul and Silke Hamann (to appear). Loanword adaptation as first-language phonological perception. To appear in: Andrea Calabrese and W. Leo Wetzels, *Loanword Phonology.* Amsterdam: John Benjamins. [ROA-975]

Boersma, Paul and Bruce Hayes (2001). Empirical Tests of the Gradual Learning Algorithm. Linguistic inquiry 32 (1), 45-86.

Boersma, Paul and David Weenink (2009). *Praat: doing phonetics by computer* (Version 5.1.07). [Computer program]. Retrieved May 12, 2009, from www.praat.org

Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English.* New York: Harper & Row.

Čermák, František (1996). Systém, funkce, forma a sémantika českých předložek. (System, function, form and meaning of Czech prepositions.) *Slovo a slovesnost* 57, 30-46.

Dickins, Tom (1998). Prepositional vocalization in contemporary Czech. *The Slavonic and East European Review* 76 (2), 201-233.

Goldsmith, John (1976). *Autosegmental Phonology.* PhD dissertation, MIT.

Hála, Bohuslav (1962). *Uvedení do fonetiky češtiny na obecně fonetickém základě*. (An introduction to the phonetics of Czech with a general phonetics background.) Prague: ČSAV.

Havránek, Bohuslav and Alois Jedlička (1981). *Česká mluvnice.* (Czech grammar.) Prague: SPN.

Hayes, Bruce (1995). *Metrical stress theory: principles and case studies*. The University of Chicago Press.

Hruška, Jiří (1976). An attempt at a linguistic characterology of prepositions in present day English in comparison with Czech. *Brno studies in English* 12: 125-144. http://www.phil.muni.cz/plonedata/wkaa/BSE/BSE_1976-12_Scan/BSE_12_05.pdf

Jakobson, Roman, Gunnar Fant and Morris Halle (1952). *Preliminaries to Speech Analysis.* Cambridge, MA.: MIT Press.

Kager, René (1995). Consequences of catalexis. In: H. van der Hulst & J. van de Weijer (eds.) *Leiden in Last: HIL Phonology Papers I*, The Hague: Holland Academic Graphics, 269-298.

Kager, René (1999). *Optimality Theory*. Cambridge: CUP.

Kenstowicz, Michael (1996). Base-identity and uniform experience: alternatives to cyclicity. In J. Durand and B. Laks (eds.), *Current trends in phonology: models and methods.* CNRS, Paris X, and University of Salford. University of Salford Publications, 363-393. [ROA-103]

Kirchner, Robert (1998). *An Effort-Based Approach to Consonant Lenition.* PhD dissertation, UCLA. [ROA-276]

Krčmová. Marie. (1984). *Fonetika a fonologie českého jazyka.* (Phonetics and phonology of Czech.) Prague: SPN.

Kučera, Henry (1961). *The phonology of Czech.* Den Haag: Mouton & Co.

Kučera, Karel (1984). K vokalizaci neslabičných předložek v současné češtině. (On the vocalization of non-syllabic prepositions in contemporary Czech.) *Naše Řeč* 57, 225-237.

Mann, Stuart Edward (1977). *Czech historical grammar*. Hamburg: Buske Verlag.

McCarthy, John (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry* 17, 207-264.

McCarthy, John (1995). Extensions of faithfulness: Rotuman revisited. Ms. University of Massachusetts, Amherst.

McCarthy, John (1998). Sympathy & phonological opacity. Ms. University of Massachusetts, Amherst. [ROA-252-0398].

McCarthy, John and Alan Prince (1986). Prosodic Morphology: Constraint Interaction and Satisfaction. Ms. University of Massachusetts, Amherst and Rutgers University. [ROA-482-1201]

McCarthy, John and Alan Prince (1993). Generalized Alignment. In Geert Booij and Jaap van Marle (eds). *Yearbook of Morphology.* Dordrecht: Kluwer. 79-153. [ROA-7-0000].

Pająk, Bożena and Eric Baković (2009). Assimilation, antigemination, and contingent optionality: the phonology of monoconsonantal proclitics in Polish. To appear in *Natural Language and Linguistic Theory*. [ROA-1034]

Palková, Zdena (1994). *Fonetika a fonologie češtiny.* (Czech phonetics and phonology.) Praha: Karolinum.

Petr, Jan (Ed.) *et al*. (1986). *Mluvnice češtiny (1). Fonetika, fonologie, morfonologie, tvoření slov*. (Czech grammar (1). Phonetics, phonology, morphophonology, word formation.) Prague: Academia.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Manuscript, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Massachusetts & Oxford: Blackwell.

Rubach, Jerzy (2000). Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31 (2), 271-317. MIT Press.

Scheer, Tobias (1996). *Une théorie de l'interaction directe entre consonnes.* (A theory of direct interaction between consonants.) PhD dissertation, Université Paris 7.

Smolensky, Paul (1997). Random rules in Universal Grammar. Talk presented at the Hopkins Optimality Theory Conference.

Smolensky, Paul, and Géraldine Legendre (2006). *The harmonic mind: From neural computation to Optimality-Theoretic grammar.* Cambridge, MA: MIT Press.

Tesar, Bruce and Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Trávníček, František (1949) *Mluvnice spisovné češtiny.* (Grammar of literary Czech.) Prague: Slovanské nakladatelství.

Ziková, Markéta (2008). *Alternace vokálů s nulou v současné češtině – laterální autosegmentální analýza*. (Vowel-zero alternations in contemporary Czech – lateral autosegmental analysis.) PhD dissertation, Masarykova Univerzita Brno.

Czech national corpus – SYN2005. *Český národní korpus - SYN2005.* Institute of the Czech national corpus, FF UK, Prague 2005. Available from: http://www.korpus.cz

Acknowledgements

I am very grateful to my supervisor Paul Boersma for all that I have been able to learn from him in the course of the past year – starting from his courses, ending in the long series of dialogues on phonology, phonetics and modeling that we have had. I would like to thank Wolfgang Kehrein for the great lot he taught me about phonology. To both of them I thank for their willingness to discuss problems whenever I needed, and for commenting on the previous version of this thesis.

The most special thanks go to my mother and to my sister, and most importantly to David.

Appendix

(1) The initial grammar.

File type = "ooTextFile"

Object class = "OTGrammar 2"

<OptimalityTheory>

0 ! leak

26 constraints

constraint [1]: "*<to> |ke|" 100 100 1

constraint [2]: "*/CC/ [_C:]" 100 100 1

constraint [3]: "*[CiCi]art" 100 100 1

constraint [4]: "MAX" 100 100 1

constraint [5]: "Align" 100 100 1

constraint [6]: "Voi" 100 100 1

constraint [7]: "*/x/ [ ]" 100 100 1

constraint [8]: "*/ / [x]" 100 100 1

constraint [9]: "MinWord" 100 100 1

constraint [10]: "*/CCC/ [7cue]" 100 100 1

constraint [11]: "*/CCCC/ [7cue]" 100 100 1

constraint [12]: "*/CCC/ [6cue]" 100 100 1

constraint [13]: "DEP" 100 100 1

constraint [14]: "*\he/O/=/S/" 100 100 1

constraint [15]: "*<to>|k|" 100 100 1

constraint [16]: "IDword" 100 100 1

constraint [17]: "IDprep" 100 100 1

constraint [18]: "*/+,-/[-,+]" 100 100 1

constraint [19]: "/\a',a/[a,A]" 100 100 1

constraint [20]: "*/OSO/" 100 100 1

constraint [21]: "*O/nucl" 100 100 1

constraint [22]: "Parse" 100 100 1

constraint [23]: "*[C?C]art" 100 100 1

constraint [24]: "*<car>|auto|" 100 100 1

constraint [25]: "*<car>|\?gauto|" 100 100 1

constraint [26]: "Stress 1st" 100 100 1

0 fixed rankings

10 tableaus

input [1]: "<to+team>" 40

"|k+ti\:fmu| /k.t\i'\:f.mu/ [ktI\:fmu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+ti\:fmu| /k.t\i'\:f.mu/ [ketI\:fmu]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+ti\:fmu| /kt\i'\:f.mu/ [ktI\:fmu]" 0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+ti\:fmu| /k\e'.ti\:f.mu/ [kti\:fmu]" 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0

"|k+ti\:fmu| /k\e'.ti\:f.mu/ [kEti\:fmu]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k.t\i'\:f.mu/ [ktI\:fmu]" 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

"|ke+ti\:fmu| /k.t\i'\:f.mu/ [ketI\:fmu]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

"|ke+ti\:fmu| /kt\i'\:f.mu/ [ktI\:fmu]" 1 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k\e'.ti\:f.mu/ [kti\:fmu]" 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k\e'.ti\:f.mu/ [kEti\:fmu]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k\e'.ti\:f.mu/ [ktI\:fmu]" 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k\e'.ti\:f.mu/ [kti\:fmU]" 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0

"|k+ti\:fmu| /kt\i'\:f.mu/ [kti\:fmU]" 0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0

"|k+ti\:fmu| /k\e'.ti\:f.mu/ [ktI\:fmu]" 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0

"|ke+ti\:fmu| /ke.t\i'\:f.mu/ [ktI\:fmu]" 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

"|ke+ti\:fmu| /ke.t\i'\:f.mu/ [kti\:fmU]" 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

"|k+ti\:fmu| /kti\:f.m\u'/ [kti\:fmU]" 0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1

"|k+ti\:fmu| /ke.t\i'\:f.mu/ [ktI\:fmu]" 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 1

"|k+ti\:fmu| /k.t\i'\:fm/ [ktI\:fm]" 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+ti\:fmu| /k.t\i'\:f.mu/ [ktI\:fm]" 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+ti\:fmu| /k.t\i'\:f.muk/ [ktI\:fmuk]" 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+ti\:fmu| /k.t\i'\:f.mu/ [ktI\:fmuk]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|ke+ti\:fmu| /k\e'.ti\:fm/ [kEti\:fm]" 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k\e'.ti\:f.mu/ [kEti\:fm]" 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k\e'.ti\:f.muk/ [kEti\:fmuk]" 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+ti\:fmu| /k\e'.ti\:f.mu/ [kEti\:fmuk]" 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|k+ti\:fmu| /kt\i'\:fm/ [ktI\:fm]" 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+ti\:fmu| /kt\i'\:f.mu/ [ktI\:fm]" 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+ti\:fmu| /kt\i'\:f.muk/ [ktI\:fmuk]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+ti\:fmu| /kt\i'\:f.mu/ [ktI\:fmuk]" 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+ti\:fmu| /k't\i'\:fm/ [ktI\:fm]" 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|k+ti\:fmu| /k't\i'\:f.mu/ [ktI\:fm]" 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|k+ti\:fmu| /k't\i'\:f.muk/ [ktI\:fmuk]" 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0

"|k+ti\:fmu| /k't\i'\:f.mu/ [ktI\:fmuk]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|k+ti\:fmu| /k't\i'\:f.mu/ [ktI\:fmu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|k+ti\:fmu| /k't\i'\:f.mu/ [ketI\:fmu]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|ke+ti\:fmu| /k't\i'\:f.mu/ [ketI\:fmu]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

"|ke+ti\:fmu| /k't\i'\:f.mu/ [ktI\:fmu]" 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

"|k+ti\:fmu| /.kt\i'\:f (m) \u'./ [ktI\:fmU]" 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|ke+ti\:fmu| /.k (e) t\i'\:f.mu./ [ketI\:fmu]" 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

input [2]: "<to+house>" 34

"|k+domu| /g.d\o'.mu/ [gdOmu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 1

"|k+domu| /k.d\o'.mu/ [gdOmu]" 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 1

"|k+domu| /k.d\o'.mu/ [kedOmu]" 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+domu| /k\e'.do.mu/ [kEdomu]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+domu| /k\e'.do.mu/ [gdomu]" 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 1

"|k+domu| /k.t\o'.mu/ [gdOmu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 2 0 0 1 0 0 0 0 1

"|k+domu| /k.t\o'.mu/ [ktOmu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 1

"|k+domu| /k.d\o'.mu/ [kdOmu]" 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|ke+domu| /g.d\o'.mu/ [gdOmu]" 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1

```
"|ke+domu| /k.d\o'.mu/ [gdOmu]"     1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1
"|ke+domu| /k.d\o'.mu/ [kedOmu]"    1 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
"|ke+domu| /k\e'.do.mu/ [kEdomu]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+domu| /k\e'.do.mu/ [gdomu]"    1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0
"|ke+domu| /k.t\o'.mu/ [gdOmu]"     1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 2 0 0 1 0 0 0 0 1
"|k+domu| /gd\o'.mu/ [gdOmu]"               0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
"|k+domu| /kd\o'.mu/ [gdOmu]"               0 0 0 0 2 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0
"|k+domu| /kd\o'.mu/ [kedOmu]"              0 0 0 0 2 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|k+domu| /kt\o'.mu/ [gdOmu]"               0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 1 0 2 0 0 0 0 0 0 0 0
"|k+domu| /kt\o'.mu/ [ktOmu]"               0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0
"|k+domu| /kd\o'.mu/ [kdOmu]"               0 0 0 0 2 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|ke+domu| /gd\o'.mu/ [gdOmu]"              1 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
"|ke+domu| /kd\o'.mu/ [gdOmu]"              1 0 0 1 2 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
"|ke+domu| /kd\o'.mu/ [kedOmu]"     1 0 0 1 2 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+domu| /kt\o'.mu/ [gdOmu]"              1 0 0 1 2 0 0 0 0 0 0 0 0 0 0 1 0 2 0 0 0 0 0 0 0 0
"|k+domu| /g'd\o'.mu/ [gdOmu]"              0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0
"|k+domu| /k'd\o'.mu/ [gdOmu]"              0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0
"|k+domu| /k'd\o'.mu/ [kedOmu]"     0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
"|k+domu| /k't\o'.mu/ [gdOmu]"              0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 2 0 0 0 1 0 0 0 0
"|ke+domu| /g'd\o'.mu/ [gdOmu]"     1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0
"|ke+domu| /k'd\o'.mu/ [gdOmu]"     1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0
"|ke+domu| /k'd\o'.mu/ [kedOmu]"    1 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+domu| /k't\o'.mu/ [gdOmu]"     1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 2 0 0 0 1 0 0 0 0
"|k+domu| /.kd\o'(m)\u'./ [kdOmU]"           0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
"|ke+domu| /.k(e)d\o'.mu./ [kedOmu]"  1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

input [3]: "<to+bike>" 20
"|k+kolu| /k.k\o'.lu/ [kekOlu]"         0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 0
"|k+kolu| /k\e'.ko.lu/ [kEkolu]"        0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0
"|k+kolu| /k.k\o'.lu/ [kkOlu]"                  0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1
"|k+kolu| /k.k\o'.lu/ [k:Olu]"          0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1
"|ke+kolu| /k\e'.ko.lu/ [kEkolu]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+kolu| /k.k\o'.lu/ [kekOlu]"        1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
"|ke+kolu| /k.k\o'.lu/ [k:Olu]"         1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
"|ke+kolu| /k.k\o'.lu/ [kkOlu]"         1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
"|k+kolu| /kk\o'.lu/ [kekOlu]"                  0 0 0 0 2 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|k+kolu| /kk\o'.lu/ [kkOlu]"           0 0 1 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|k+kolu| /kk\o'.lu/ [k:Olu]"           0 1 0 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|ke+kolu| /kk\o'.lu/ [kekOlu]"         1 0 0 1 2 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+kolu| /kk\o'.lu/ [k:Olu]"                  1 1 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+kolu| /kk\o'.lu/ [kkOlu]"                  1 0 1 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|k+kolu| /k'k\o'.lu/ [kekOlu]"         0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
"|k+kolu| /k'k\o'.lu/ [kkOlu]"                  0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
"|k+kolu| /k'k\o'.lu/ [k:Olu]"          0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
```

"|ke+kolu| /k'k\o'.lu/ [kekOlu]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+kolu| /k'k\o'.lu/ [k:Olu]" 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+kolu| /k'k\o'.lu/ [kkOlu]" 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

input [4]: "<to+goal>" 26
"|k+go\:flu| /k.g\o'\:f.lu/ [kegO\:flu]" 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1
"|k+go\:flu| /k\e'.go\:f.lu/ [kEgo\:flu]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0
"|k+go\:flu| /g\o'\:f.lu/ [kegO\:flu]" 0 0 0 1 0 0 0 2 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|k+go\:flu| /g.g\o'\:f.lu/ [ggO\:flu]" 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 1
"|k+go\:flu| /k.g\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 1 0 1 0 0 0 0 0
"|ke+go\:flu| /k\e'.go\:f.lu/ [kEgo\:flu]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+go\:flu| /k.g\o'\:f.lu/ [kegO\:flu]" 1 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
"|ke+go\:flu| /g\o'\:f.lu/ [kegO\:flu]" 1 0 0 2 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+go\:flu| /g.g\o'\:f.lu/ [ggO\:flu]" 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 1
"|ke+go\:flu| /k.g\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 1 1 0 1 0 0 0 0 1
"|k+go\:flu| /g.g\o'\:f.lu/ [ggO\:flu]" 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 1
"|k+go\:flu| /k.g\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 1
"|ke+go\:flu| /g.g\o'\:f.lu/ [ggO\:flu]" 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 1
"|ke+go\:flu| /k.g\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 1
"|k+go\:flu| /kg\o'\:f.lu/ [kegO\:flu]" 0 0 0 0 2 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|k+go\:flu| /gg\o'\:f.lu/ [ggO\:flu]" 0 0 1 0 2 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
"|k+go\:flu| /kg\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 0 2 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0
"|ke+go\:flu| /kg\o'\:f.lu/ [kegO\:flu]" 1 0 0 1 2 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+go\:flu| /gg\o'\:f.lu/ [ggO\:flu]" 0 0 1 1 2 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
"|ke+go\:flu| /kg\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 1 2 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0
"|k+go\:flu| /k'g\o'\:f.lu/ [kegO\:flu]" 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
"|k+go\:flu| /g'g\o'\:f.lu/ [ggO\:flu]" 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0
"|k+go\:flu| /k'g\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0
"|ke+go\:flu| /k'g\o'\:f.lu/ [kegO\:flu]" 1 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+go\:flu| /g'g\o'\:f.lu/ [ggO\:flu]" 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0
"|ke+go\:flu| /k'g\o'\:f.lu/ [g\:fO\:flu]" 0 1 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0

input [5]: "<to+car>" 44
"|k+autu| /k.\a'u.tu/ [k\?gAUtu]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1
"|k+autu| /k.\?g\a'u.tu/ [k\?gAUtu]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 1 0 1
"|k+autu| /k.\?g\a'u.tu/ [kAUtu]" 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 1 0 1
"|k+autu| /k\a'u.tu/ [k\?gAUtu]" 0 0 0 0 2 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0
"|k+autu| /k\a'u.tu/ [kAUtu]" 0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0
"|k+autu| /ke.\?g\a'u.tu/ [k\?gautu]" 0 0 0 0 2 0 1 0 0 0 0 0 2 0 1 0 0 0 1 0 0 0 1 1 0 0
"|k+autu| /ke.\a'u.tu/ [kEautu]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0
"|k+autu| /k.\a'u.tu/ [kAUtu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0 1
"|ke+autu| /k.\a'u.tu/ [k\?gAUtu]" 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 1
"|ke+autu| /k.\?g\a'u.tu/ [k\?gAUtu]" 1 0 0 1 2 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 1 0 1
"|ke+autu| /k.\?g\a'u.tu/ [kAUtu]" 1 0 0 1 2 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 1

"|ke+autu| /k\a'u.tu/ [k\?gAUtu]" 1 0 0 1 2 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0
"|ke+autu| /ke.\?g\a'u.tu/ [k\?gautu]" 1 0 0 0 1 0 2 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 1 0 0
"|ke+autu| /ke.\?g\a'u.tu/ [kautu]" 1 0 0 0 1 0 2 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0
"|ke+autu| /ke.\a'u.tu/ [kEautu]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
"|ke+autu| /k\a'u.tu/ [kAUtu]" 1 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
"|ke+autu| /k.\a'u.tu/ [kAUtu]" 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1
"|k+autu| /k\?g\a'u.tu/ [k\?gAUtu]" 0 0 0 0 2 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 1 0 0
"|ke+autu| /k\?g\a'u.tu/ [k\?gAUtu]" 1 0 0 1 2 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0
"|k+\?gautu| /k\?g\a'u.tu/ [k\?gAUtu]" 0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0
"|k+\?gautu| /k\?g\a'u.tu/ [kAUtu]" 0 0 0 0 2 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0
"|k+\?gautu| /k\a'u.tu/ [kAUtu]" 0 0 0 1 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0
"|ke+\?gautu| /k\?g\a'u.tu/ [k\?gAUtu]" 1 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0
"|ke+\?gautu| /k\?g\a'u.tu/ [kAUtu]" 1 0 0 1 2 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
"|ke+\?gautu| /k\a'u.tu/ [kAUtu]" 1 0 0 2 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
"|k+\?gautu| /k.\?g\a'u.tu/ [k\?gAUtu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 0 1 1
"|k+\?gautu| /k.\?g\a'u.tu/ [kAUtu]" 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 1
"|k+\?gautu| /k.\a'u.tu/ [kAUtu]" 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 1
"|ke+\?gautu| /k.\?g\a'u.tu/ [k\?gAUtu]" 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
"|ke+\?gautu| /k.\?g\a'u.tu/ [kAUtu]" 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1
"|ke+\?gautu| /k.\a'u.tu/ [kAUtu]" 1 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1
"|k+\?gautu| /k'\?g\a'u.tu/ [k\?gAUtu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 1 0
"|k+\?gautu| /k'\?g\a'u.tu/ [kAUtu]" 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0
"|k+\?gautu| /k'\a'u.tu/ [kAUtu]" 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0
"|ke+\?gautu| /k'\?g\a'u.tu/ [k\?gAUtu]" 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0
"|ke+\?gautu| /k'\?g\a'u.tu/ [kAUtu]" 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0
"|ke+\?gautu| /k'\a'u.tu/ [kAUtu]" 1 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0
"|k+autu| /k'\a'u.tu/ [k\?gAUtu]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 1 0 0
"|k+autu| /k'\?g\a'u.tu/ [k\?gAUtu]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 1 1 0 0
"|k+autu| /k'\?g\a'u.tu/ [kAUtu]" 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 1 0 0
"|k+autu| /k'\a'u.tu/ [kAUtu]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0
"|ke+autu| /k'\a'u.tu/ [k\?gAUtu]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
"|ke+autu| /k'\?g\a'u.tu/ [k\?gAUtu]" 1 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 1 1 0 0
"|ke+autu| /k'\a'u.tu/ [kAUtu]" 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0

input [6]: "<to+chink>" 16
"|k+\shkvi\:f\r<e| /k.\shkv\i'\:f.\r<e/ [ke\shkvi\:f\r<e]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1
"|k+\shkvi\:f\r<e| /k.\shkv\i'\:f.\r<e/ [k\shkvi\:f\r<e]" 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1
"|k+\shkvi\:f\r<e| /k\e'.\shkvi\:f.\r<e/ [kE\shkvi\:f\r<e]" 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
"|k+\shkvi\:f\r<e| /k\e'.\shkvi\:f.\r<e/ [k\shkvi\:f\r<e]" 0 0 0 0 1 0 1 0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0
"|ke+\shkvi\:f\r<e| /k.\shkv\i'\:f.\r<e/ [ke\shkvi\:f\r<e]" 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0
"|ke+\shkvi\:f\r<e| /k\e'.\shkvi\:f.\r<e/ [kE\shkvi\:f\r<e]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+\shkvi\:f\r<e| /k.\shkv\i'\:f.\r<e/ [k\shkvi\:f\r<e]" 1 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
"|ke+\shkvi\:f\r<e| /k\e'.\shkvi\:f.\r<e/ [k\shkvi\:f\r<e]" 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
"|k+\shkvi\:f\r<e| /k\shkv\i'\:f.\r<e/ [ke\shkvi\:f\r<e]" 0 0 0 0 2 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

"|k+\shkvi\:f\r<e| /k\shkv\i'\:f.\r<e/ [k\shkvi\:f\r<e]" 0 0 0 0 2 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
"|ke+\shkvi\:f\r<e| /k\shkv\i'\:f.\r<e/ [k\shkvi\:f\r<e]" 1 0 0 1 2 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+\shkvi\:f\r<e| /k\shkv\i'\:f.\r<e/ [ke\shkvi\:f\r<e]" 1 0 0 1 2 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|k+\shkvi\:f\r<e| /k'\shkv\i'\:f.\r<e/ [ke\shkvi\:f\r<e]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
"|k+\shkvi\:f\r<e| /k'\shkv\i'\:f.\r<e/ [k\shkvi\:f\r<e]" 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+\shkvi\:f\r<e| /k'\shkv\i'\:f.\r<e/ [k\shkvi\:f\r<e]" 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+\shkvi\:f\r<e| /k'\shkv\i'\:f.\r<e/ [ke\shkvi\:f\r<e]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

input [7]: "<to+dog1>" 16
"|k+psu| /k.ps\u'/ [kepsU]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1
"|k+psu| /k.ps\u'/ [kpsU]" 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1
"|k+psu| /k\e'.psu/ [kEpsu]" 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
"|k+psu| /k\e'.psu/ [kpsu]" 0 0 0 0 1 0 1 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0
"|ke+psu| /k.ps\u'/ [kepsU]" 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1
"|ke+psu| /k.ps\u'/ [kpsU]" 1 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1
"|ke+psu| /k\e'.psu/ [kEpsu]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+psu| /k\e'.psu/ [kpsu]" 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
"|k+psu| /kps\u'/ [kepsU]" 0 0 0 0 2 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
"|k+psu| /kps\u'/ [kpsU]" 0 0 0 0 2 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+psu| /kps\u'/ [kepsU]" 1 0 0 1 2 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+psu| /kps\u'/ [kpsU]" 1 0 0 1 2 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|k+psu| /k'ps\u'/ [kepsU]" 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
"|k+psu| /k'ps\u'/ [kpsU]" 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+psu| /k'ps\u'/ [kepsU]" 1 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+psu| /k'ps\u'/ [kpsU]" 1 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

input [8]: "<to+dog2>" 16
"|k+psov\ic| /k.ps\o'.v\ic/ [kepsOv\ic]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1
"|k+psov\ic| /k.ps\o'.v\ic/ [kpsOv\ic]" 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1
"|k+psov\ic| /k\e'.pso.v\ic/ [kEpsov\ic]" 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
"|k+psov\ic| /k\e'.pso.v\ic/ [kpsov\ic]" 0 0 0 0 1 0 1 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0
"|ke+psov\ic| /k.ps\o'.v\ic/ [kepsOv\ic]" 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1
"|ke+psov\ic| /k\e'.pso.v\ic/ [kEpsov\ic]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+psov\ic| /k\e'.pso.v\ic/ [kpsov\ic]" 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
"|ke+psov\ic| /k.ps\o'.v\ic/ [kpsOv\ic]" 1 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1
"|k+psov\ic| /kps\o'.v\ic/ [kepsOv\ic]" 0 0 0 0 2 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
"|k+psov\ic| /kps\o'.v\ic/ [kpsOv\ic]" 0 0 0 0 2 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+psov\ic| /kps\o'.v\ic/ [kepsOv\ic]" 1 0 0 1 2 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|ke+psov\ic| /kps\o'.v\ic/ [kpsOv\ic]" 1 0 0 1 2 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
"|k+psov\ic| /k'ps\o'.v\ic/ [kepsOv\ic]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
"|k+psov\ic| /k'ps\o'.v\ic/ [kpsOv\ic]" 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+psov\ic| /k'ps\o'.v\ic/ [kepsOv\ic]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
"|ke+psov\ic| /k'ps\o'.v\ic/ [kpsOv\ic]" 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

input [9]: "<to+fence>" 16

"|k+plotu| /k.pl\o'.tu/ [keplOtu]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+plotu| /k\e'.plo.tu/ [kEplotu]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+plotu| /k.pl\o'.tu/ [kplOtu]" 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+plotu| /k\e'.plo.tu/ [kplotu]" 0 0 0 0 1 0 1 0 0 1 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0

"|ke+plotu| /k.pl\o'.tu/ [keplOtu]" 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

"|ke+plotu| /k\e'.plo.tu/ [kEplotu]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+plotu| /k.pl\o'.tu/ [kplOtu]" 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

"|ke+plotu| /k\e'.plo.tu/ [kplotu]" 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0

"|k+plotu| /kpl\o'.tu/ [keplOtu]" 0 0 0 0 2 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+plotu| /kpl\o'.tu/ [kplOtu]" 0 0 0 0 2 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0

"|ke+plotu| /kpl\o'.tu/ [keplOtu]" 1 0 0 1 2 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+plotu| /kpl\o'.tu/ [kplOtu]" 1 0 0 1 2 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|k+plotu| /k'pl\o'.tu/ [keplOtu]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|k+plotu| /k'pl\o'.tu/ [kplOtu]" 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0

"|ke+plotu| /k'pl\o'.tu/ [keplOtu]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

"|ke+plotu| /k'pl\o'.tu/ [kplOtu]" 1 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

input [10]: "<to+mercury>" 20

"|k+rtuc\ic| /k.rt\u'.c\ic/ [kertUc\ic]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+rtuc\ic| /k\e'.rtuc\ic/ [kErtuc\ic]" 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

"|k+rtuc\ic| /k.rt\u'.c\ic/ [krtUc\ic]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1

"|k+rtuc\ic| /\k'.rtu.c\ic/ [Krtuc\ic]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0

"|k+rtuc\ic| /k\e'.rtuc\ic/ [krtuc\ic]" 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0

"|k+rtuc\ic| /k\r'.tuc\ic/ [kRtuc\ic]" 0 0 0 0 2 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0

"|k+rtuc\ic| /k\r'.tuc\ic/ [krtuc\ic]" 0 0 0 0 2 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 1 0 0 0 0 0

"|ke+rtuc\ic| /k.rt\u'.c\ic/ [kertUc\ic]" 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

"|ke+rtuc\ic| /k\e'.rtuc\ic/ [kErtuc\ic]" 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

"|ke+rtuc\ic| /k.rt\u'.c\ic/ [krtUc\ic]" 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

"|ke+rtuc\ic| /k\e'.rtuc\ic/ [krtuc\ic]" 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0

"|ke+rtuc\ic| /k\r'.tuc\ic/ [kRtuc\ic]" 1 0 0 1 2 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

"|k+rtuc\ic| /krt\u'.c\ic/ [kertUc\ic]" 0 0 0 0 2 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0

"\he|k+rtuc\ic| /krt\u'.c\ic/ [krtUc\ic]" 0 0 0 0 2 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0

"|ke+rtuc\ic| /krt\u'.c\ic/ [kertUc\ic]" 1 0 0 1 2 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0

"|ke+rtuc\ic| /krt\u'.c\ic/ [krtUc\ic]" 1 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0

"|k+rtuc\ic| /k'rt\u'.c\ic/ [kertUc\ic]" 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0

"|k+rtuc\ic| /k'rt\u'.c\ic/ [krtUc\ic]" 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0

"|ke+rtuc\ic| /k'rt\u'.c\ic/ [kertUc\ic]" 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0

"|ke+rtuc\ic| /k'rt\u'.c\ic/ [krtUc\ic]" 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0

(2) The adult output distribution.

"ooTextFile"

"Distributions"

1 column

"Czech"

15 rows

"[ktI\:fmu]"	165
"[gdOmu]"	165
"[k\?gAUtu]"	47
"[kAUtu]"	3
"[kEkolu]"	10
"[kEgo\:flu]"	10
"[kEpsu]"	8
"[kplOtu]"	27
"[kpsOv\ic]"	15
"[k\shkvi\:f\r<e]"	3
"[krtUc\ic]"	2
"[kEpsov\ic]"	15
"[kE\shkvi\:f\r<e]"	27
"[kEplotu]"	3
"[kErtuc\ic]"	6

(3) The script for learning.

```
distrName$ = "distr_0625"
grammarName$ = "grammar0625"
for i to 10
	Read from file... 'distrName$'.txt
	Read from file... 'grammarName$'.txt
	select Distributions 'distrName$'
	plus OTGrammar 'grammarName$'
	Learn from partial outputs... 1 2 "Weighted all" 1 100000 0.1 4 0.1 yes 1 0
	select OTGrammar 'grammarName$'
	To PairDistribution... 10000 2
	To Table
	Extract rows where column (number)... weight "greater than or equal to" 100
	Write to table file... results\learner'i'.txt
	select Distributions 'distrName$'
	plus PairDistribution 'grammarName$'_out
	Remove
	select OTGrammar 'grammarName$'
	Rename... grammar'i'
	Write to text file... results\grammar'i'.txt
endfor
```