ROCZNIKI HUMANISTYCZNE
Tom LXXIII, zeszyt 11 – 2025
ZESZYT SPECJALNY / SPECIAL ISSUE
DOI: https://doi.org/10.18290/rh257311.1s

ANNA BLOCH-ROZMEJ

THE PHONOTACTICS OF WORD EDGES IN LOVARI ROMANI

Abstract. This paper addresses the question of phonotactic conditions governing the structure of North West Lovari word edges. The problem is analysed within the framework of Government Phonology. Our discussion is intended to shed new light on the aspects of the system that have not been subject to analysis within any non-derivational framework yet. The present study is corpus-based, and the relevant data come from the Romani Morpho-Syntax database (Romani Project).

Keywords: Romani; phonology; phonotactics; clusters; corpus; representation; principle; constituent; consonant; parameter

FONOTAKTYKA BRZEGÓW DOMEN FONOLOGICZNYCH W LOVARI ROMANI

Abstrakt. Artykuł porusza problem fonotaktyki domen fonologicznych w języku Lovari Romani. Autorka bada, jakie zbitki spółgłoskowe są dozwolone na początku oraz na końcu wyrazów w języku romskim oraz jakie struktury fonologiczne manifestują się fonetycznie za pomocą tych grup segmentalnych. Przedstawiona analiza jest pierwszą analizą niederywacyjną fonotaktyki języka Lovari Romani. Modelem teoretycznym zastosowanym w analizie jest model fonologii rządu, zaś dane pochodzą z korpusu Romani Project – RMS Database.

Słowa kluczowe: język romski; fonologia; fonotaktyka; korpus; reprezentacja; zbitki spółgłoskowe; parametr

ANNA BLOCH-ROZMEJ, PhD, Dr Litt., Assistant Professor at the John Paul II Catholic University of Lublin, Institute of Linguistics, Department of Applied Linguistics; correspondence address: Instytut Językoznawstwa KUL, Al. Racławickie 14, 20-950 Lublin, Poland; e-mail: anna.bloch-rozmej@kul.pl; ORCID: https://orcid.org/0000-0001-8675-1726.

INTRODUCTION

The aim of this paper is to investigate phonotactic conditions of word structure in Lovari Romani. More specifically, we shall look into the inventory of word-initial and word-final consonant clusters allowed by this system and try to establish what lexical representations they manifest. The analysis will be couched within the classical framework of Government Phonology (henceforth GP) (Bloch-Rozmej, 2008; Charette, 1991; Cyran, 2003; Harris, 1994; Kaye et al., 1985, 1990), which assumes that concrete phonetic effects are strictly determined by phonological representations, universal principles and language-specific parameters. The present discussion is meant to broaden our understanding of the organization of the phonological system of Lovari Romani by proposing the first non-derivational analysis of the language's consonantal clusters attested at the edges of words. We shall identify phonological constituent types that organise word-initial and word-final melodic material in this dialect.

The dialect of Romani that we analyse in this paper is that of the North West variety of Lovari, i.e. the Hungarised Vlax. Romani is a language spoken in most parts of Europe, both Americas and Turkey, though the main region of its use is the Balkan Peninsula. Matras (2002) classifies Romani as an Indo-Aryan language. As noted in Wagner (2012, p. 4), many lexical and grammatical properties of the language suggest its Indian origin. Boretzky (2003, p. 1) points out that the speakers of Vlax live in Ukraine, South Russia, the Balkan countries, Greece, Turkey, Hungary, and in Slovakia. Within the Vlax dialect, we can distinguish a Northern (mainly Romanian or non-Balkan) group and a Southern one (in the Balkans outside Romania) (Boretzky, 2003, p. 90). The Northern group can be divided into a Hungarian dominated (Western) branch and a less homogeneous Eastern group (Boretzky, 2003, p. 97). The speakers of the former dialect, who call themselves "Lovára" (SG Lovári), inhabit what used to be the Austrian-Hungarian monarchy as well as neighbouring regions including Germany and Southern Poland. The features of the dialect presented in the subsequent sections pertain to the language variety whose users live primarily in Czech and Slovak speaking environments. According to the Statistical Office of the Slovak Republic *Štatistický úrad Slovenskej Republiky* (2001), 0.9% of Slovakia's population declared that their mother tongue was Romani. Lakatošová and Šebková (2004, p. 2) point out that speakers of North West Lovari Romani live in Bratislava, Galanta, Lučenec, Nitra, Nové Zámky, Sered' and Šal'a. Lovari Romani occurs mainly in oral speech and its written records are rather scarce, being mainly dictations of spoken language. Thus,

in Section 2, the most characteristic features of the Lovari Romani sound system will be discussed.

The article is organised as follows. We start by discussing the relevant facts concerning the phonological system of Lovari Romani established so far and available in the existing sources of which Wagner (2012) is of greatest importance. This presentation comprises phonetic characteristics of Lovari consonants and the inventory of the vocalic sounds. We shall also indicate a number of processes that these segments can undergo in Lovari words. All the examples addressed in this article come from the project Romani Morpho-Syntax Database (RMS). The analysis involves investigating the structure of 953 sample phrases of Lovari Romani spoken in West Slovakia with a view to compiling a set of the attested consonant clusters. This stage will be succeeded by a GP analysis of Lovari Romani word structure, as a result of which the language's phonotactic conditions pertaining to the edges of phonological domains will be proposed. We shall conclude with a brief typological note—a comparison of the phonotactic conditions of Lovari Romani with those of Polish, English and Irish.

1. MATERIALS AND METHODS

The analysis proposed in this paper is strongly based on, on the one hand, the existing descriptions of the phonology and phonetics of Lovari Romani presented in Wagner (2012) and, on the other, the corpus data obtained from the Romani Project—the RMS Database.[1] As explained on its webpage, "the Romani Project [is] a cluster of academic research activities based at the School of Arts, Languages and Cultures at the University of Manchester". The Romani Morpho-Syntax Database, which is part of the project, was conceptualized by Yaron Matras and Viktor Elšík, and designed and developed by Christopher White. The current list of the available samples contains 120 Romani dialects spoken in different countries all over the world. The data for this study come from the dialect of West Slovak (sample ref. SK-016). More precisely, the dialect is labelled as West Slovak (with transcriptions). The search-engine devised for the database makes it possible to retrieve from the selected sample phrases, transcriptions, some phonological information about the dialect such as features of word-stress, inventory of vowels, consonants

[1] Available at http://kratylos.org/~raphael/romani/atmanchester/projects/rmsdatabase.html.

(sound change), sound system and morpho-phonology as well as complex morpho-syntactic features of the dialect. The phonological characteristics, though incomplete, help us interpret the data in some problematic cases. We are also able to browse the lexico-phonetic features section of the database for useful information concerning consonant-related phenomena, such as cluster simplification, initial modifications or other sound changes attested in the selected dialect. The sample that was crucial for our analysis contained 953 phrases of Romani, each accompanied by its English translation and audio recording of its pronunciation. The sample contains examples illustrating consonant clusters attested at the edges of Lovari Romani words that will be closely examined in the forthcoming analysis. The other Romani words appearing in this article have been adopted from Wagner (2012). In cases where the meaning of a given lexical form was not quite clear from the RMS Database, we consulted the following online websites:

– the ROMLEX Lexical Database (romani.uni-graz.at/romlex),
– the Angloromani dictionary (kratylos.org/~raphael/romani/angloromani/dictionary.html)
– Glosbe: Słownik angielsko – Angloromani (glosbe.com/en/rme).

The significance of Wagner's (2012) work derives from its being a reliable source of language data because it is based on fieldwork recordings and some written texts of the North West variety of Lovari. The book also contains a description of the dialect's phonetics, phonology and sound alteration. Thus, in certain problematic cases, the data from the Romani database were counterchecked against Wagner's work.

The present analysis involves examining the internal domain structure of words found in the 953 Romani phrases that are used in the West-Slovak variety of the language and listed in the RMS database. The lexical items were analysed in terms of the consonantal content of the word-initial and word-final clusters with a view to discovering the phonological structures they derive from. Our analysis uses the tools offered by the framework of Government Phonology, which represents the non-derivational approach to phonological processing. We employ the classical version of the model (Harris, 1994; Kaye et al., 1985, 1990) since our analysis is data-based and we wish to avoid unnecessary theory-related complications. More specifically, we investigate which phonological representations underlie the phonetic manifestations of consonant clusters occurring word-initially and word-finally in the pronunciation of Lovari Romani words. Our aim is thus explanatory rather than merely descriptive as, within GP, phonological representation is perceived as the ul-

timate source of phonological events. According to the model adopted here, the syllabification of phonetic segments depends both on the universal constituent structure conditions and language-specific parameters employed by a given system, on the one hand, and the internal melodic composition of the segments themselves, on the other. Therefore, our discussion will proceed as follows. First, a brief outline of the Lovari Romani consonants and vowels will be presented. Secondly, we shall pinpoint the major consonant-related processes of the language and define the relevant theoretical assumptions of Government Phonology connected with syllable structure. Subsequently, the data-based analysis will be supplied followed by a brief typological note whereby Lovari Romani parameters determining cluster content of word-initial and word-final syllabic constituents will be compared to those of English, Polish and Irish.

2. THE SOUNDS OF LOVARI ROMANI

2.1 CONSONANTS

Wagner (2012, p. 22) argues that the consonant inventory of North West Lovari Romani comprises the following segments:

(1) *Lovari consonants according to class*

1. Plosives: [p], [t], [k], [b], [d], [g], [c], [ɟ], [ʔ] (e.g. *parno* ‘white, bright/masc.’, *tajsa* ‘tomorrow’, *kafa* ‘coffee’, [‘ʔevropa]/ [‘ʔevropa] ‘Europe’)
2. Vibrants: [r], [ɾ] and [ɾ̊]; (e.g. *řoj* ‘spoon’, *šukar* ‘nicer’)
3. Fricatives: [f], [v], [χ], [h], [ʃ], [ʃ], [s], [z] (e.g. *xarno* ‘short’, *hazd* ‘lift’, *sar* ‘how’, *zuralo* ‘strong’, *šax* ‘cabbage’, *užo* ‘clean’)
4. Affricates: [t͡s], [t͡ʃ] (e.g. *čor* ‘poor’, *dzar* ‘body hair’)
5. Nasals: [n], [ŋ], [m], [ɲ] (e.g. *an*[n]*ro* ‘egg’, *m*[m]*undar* ‘kill’, [khaŋ’ge:ri] ‘church’).
6. Approximants: [j], [l], [ʎ] (e.g. *khajni* ‘hen’, *devel* ‘god’)

The voiceless plosives [p, t, k] also have their aspirated variants ([p^{h}], [t^{h}], [k^{h}]), as in *phen* ‘say’, *akh* ‘eye’, or *athar* ‘deceive’. The system of Lovari obstruents, with a couple of exceptions, seems to be pretty symmetrical in terms of their voicing specifications. Their classification in terms of the place of articulation is as follows. The dialect possesses labials [p], [b], [ph], [m],

labio-dentals [f], [v], alveolars [t], [d], [th], [n], [r], [s], [z], [t͡s], a velarised alveolar-lateral [l], post-alveolars [ʃ], [J], [t͡ʃ], palatals [c], [ɟ], [ɲ], [ɾ̟], [r̝̊], [j], a palatal-lateral ([ʎ]), velars [k], [g], [kh], [ŋ], a uvular ([χ], and the glottals [h], [ʔ] (Wagner, 2012, p. 23).

Wagner (2012) observes that in certain contexts, consonants in North West Lovari Romani can undergo germination. This seems to be a cross-boundary phenomenon in the speech of Lovaris living in the Slovak environment which is more heavily influenced by Hungarian. Examples illustrating this development could be *phe<u>nn</u>a* 'they will say', *ža<u>nn</u>a* 'they will know', or *dikhe<u>l l</u>a* 'she sees it', where stem-final sonorants are directly followed by morpheme-initial identical consonants. Additionally, consonantal geminates can be used to express emphasis, as in *pha<u>rr</u>adi* 'bitch (fem.)' and *čo<u>rr</u>o* 'idiot'.

2.2 VOWELS

As noted in Wagner (2012), North West Lovari Romani possesses vowel segments that can be classified as front ([e], [ɛ], [i]), central ([a]) and back ([o], [u]). [i] and [u] are high, [e] and [o] are mid, whereas [a] and [ɛ] are open low vowels. What is significant here is that the low open [ɔ] is an alternative realization of [o], similar to the [ɛ]/[e] pair, the first member of the former pair occurring in open syllables for emphasis (e.g. ['kɔrɔni] ~ ['koroni]) and [ɛ] being lowered before velars and [r] (e.g. [khɛr] ~ [kher]). These changes, however, have no distinctive character. In the system of Lovari Romani, the back vowels are also rounded, while all the other vocalic expressions are unrounded. Importantly, the language exhibits phonetic oppositions in terms of vowel length. The set of long vowels includes [e:], [ɛ:], [a:], [i:], [o:], [u:]. Wagner (2012, p. 23) argues that "all vowels appear in two forms, a short and a long one, realised by a shorter or longer time before the switch to the next sign." This statement, however, does not sound too clear and to the lack of the examples of any contrastive pairs, it is difficult to establish whether vowel length is indeed distinctive in the system of Lovari. Still, it seems more likely that the system does not treat the short/long distinction as phonological since, as Wagner maintains, the long variants are used primarily for emphasis (e.g., [hɛ:j] 'hey!' or [bɐ:'rij i ʃej] 'She's big, that girl!'). The author also speaks of the existence of two diphthongs in the language, namely [ei̯] and [ou̯]. These, however, are alternative realizations of long vowels. Some examples illustrating the occurrence of long (or rather lengthened) vowels and diphthongs in

Lovari Romani, as transcribed in Wagner (2012, p. 24), are given below, in (2a) and (2b) respectively.

(2) a. *long vowels* b. *diphthongs*

[‘r̝i:sko] ‘schnitzel’	[khaŋ’ge̯i̯ri]‘church’ (besides [khaŋ’ge:ri])
[t͡sɛ:’dɛ:t͡ʃko] ‘CD’	[‘do̯u̯za] (besides [‘do:za])
[bal’ko:ni] ‘balcony’	
[pe:rma] ‘on me’	[‘pe̯i̯rma] ‘on me’
[jɔ:] ‘of course’	

The RMS Database also mentions the existence of schwa in the West Slovak variety of Romani. The vowel can be realised in unstressed word positions. Since Lovari vowels are not our major concern in this paper, we leave the vowel-related problems unresolved as more radical decisions concerning this issue would certainly require a lot more detailed investigation.

3. PHONOLOGICAL PROCESSES

Before turning to the analysis of Lovari phonotactics, we shall briefly outline the most important process types attested in this system. There are a number of changes that can target Lovari consonants. In particular, the voiced segments are likely to undergo devoicing in the word-final context and before other voiceless sounds. Consider the items listed in (3) below (Wagner, 2012, p. 28).

(3) *Consonant devoicing*

/drab/	>	[drap]	‘medicine’
/thud/	>	[thut]	‘milk’
/pánž/	>	[pa:nʃ]	‘five’
/phuv/	>	[phuf]	‘earth, soil’
/rig/	>	[rik]	‘side’
/luluď/	>	[‘luluc]	‘flower’

A consonant that is also targeted by devoicing is the Lovari /ř/. Consequently, two variants of this sound can be heard in the pronunciation of Lovari words: voiced [r̝] and voiceless [r̝̊]. The latter occurs at the end of words and before voiceless consonants.

Another development typically found in this dialect is nasal assimilation whereby the alveolar nasal [n] changes into a velar sound when followed by a

velar consonant. In fact [n] and [ŋ] are in complementary distribution. Some examples are provided in (4) (Wagner, 2012, p. 28).

(4) *Nasal assimilation*

/šing/	>	[ʃiŋg] or [ʃiŋk]	‘horn’
/inke/	>	[ʔiŋke]	‘still’
/šavourenge/	>	[ʃavo̯u̯’reŋge]	‘to the children’
/romengi/	>	[ro’meŋgi]	‘Romani’

The first example in the table illustrates this place assimilation in the word-internal context, whereas the last two depict the change of [n] into its velar counterpart at the end of the stem when followed by another velar-initial form. Lovari Romani also exhibits the loss of the glottal stop after a consonantal segment inside the speech stream. In fact, this plosive is realised only in the initial context before a vowel, irrespective of any boundaries, as in /aď in/ > [ʔa’ɟin] ‘honey’, /ejfta/ > [‘ʔei̯fta] ‘seven’, /e/ > [ʔe], /ingrel/ [ʔiŋ’grel] ‘carry’. The glottal stop is deprived of any contrastive potential and thus is not regarded to be an independent phoneme of this dialect.

The process of palatalisation appears to modify alveolar consonantal segments in this language, namely plosives, the alveolar [n] and [l] in cross-morphemic contexts. In the case of [n], the phonetic output is a palatal nasal, whereas in the case of the lateral, [j] is realised. Consider the examples below.

(5) *Palatalisation*

aď i[ɲ]atar	‘from honey’
phurd’asa	‘over the bridge’
mojange	‘from wine’
cigň[ɲ]a	‘small’
žuvja/žuvl’a	‘women’
cigň[ɲ]ol	‘become small, shrivel’

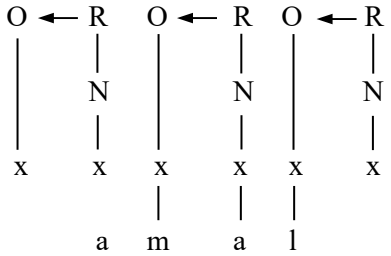
The last consonant-related development described in Wagner (2012) is de-aspiration whose sole target is [kh]. The process applies in the word-final context or word-internally when a stem-final aspirated velar is followed by another consonant. This change can be attested in such words as /dikh/ > [dik] ‘see/IMP’, or /jakh/ > [jak] ‘eye’ for instance. The focus of the subsequent section will be syllable structure of North West Lovari Romani seen through the optic of the model adopted in this study, i.e. Government Phonology.

4. THE SYLLABLE: THEORETICAL ASSUMPTIONS FOR FURTHER ANALYSIS

Traditionally, a syllable is conceived of as a central vocalic nucleus surrounded by optional consonants on the right and on the left whose sequential arrangements in the onset and the coda have to respect the Sonority Sequencing Principle (Harris, 1994, p. 56; Selkirk, 1984, p. 116; Vennemann, 1988). The concept of the syllable has no formal status within the framework of Government Phonology and the representation of the word domain has the form of the onset-nucleus (ON) sequences. The third phonological constituent recognised by the model is that of a rhyme headed by the nucleus and optionally dominating a single consonant to the right of the vowel. In more recent versions of the theory, lexical structure is reduced to CV sequences, i.e. Onset-Nucleus/Rhyme strings (e.g. Scheer, 2004). Nevertheless, whichever offshoot of Government Phonology is adopted for analysis, no coda constituent will ever be recognised. Thus, word-medial consonant clusters must necessarily be split between neighbouring OR domains. The model we choose to employ in our forthcoming analysis of the Lovari Romani data is classical Government Phonology as defined in Harris (1994). Here, a phonological word has the O-R structure where the rhyme is headed by a vowel/diphthong, which can be either short or long, while the onset is allowed to dominate maximally two consonants. The skeletal positions to which the consonants are attached have to be involved in a left-headed governing relation.[2] Also, in the case of a complex nuclear structure, yet still maximally binary branching, a similar governing relation has to be established. Such a requirement imposes rather stringent conditions on the choice of potential governors and governees, both in the case of branching nuclei and branching onsets. Consequently, potential onset governees are [r, l, j, w] since these are the least complex consonantal segments. This issue will be dealt with in more detail later in this section. Below in Figure 1 we illustrate the GP-based representation of a simple Lovari word such as *amal* 'friend'.

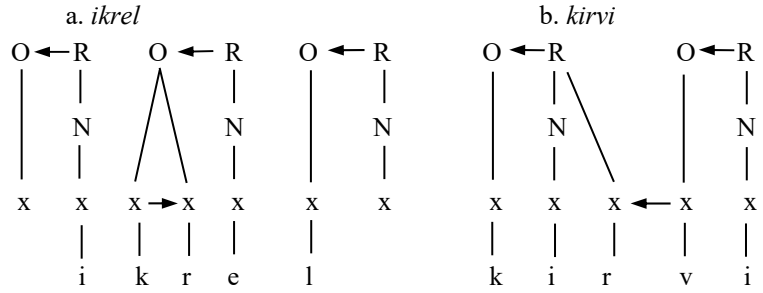
[2] Within a phonological domain, its head, the stressed nucleus, is the source of licensing potential for all the remaining skeletal positions. All these positions have to be licensed save the head itself which is licensed at higher levels of phonological hierarchy. Government is a more restrictive form of licensing, imposing more stringent conditions on the governee.

Figure 1. The structure of *amal*



As mentioned above, and depicted in Figure 1, the representation is structured as a string of OR domains which constitute governing domains (indicated by the arrow). In such structures, the onset position is universally governed by the head of the following rhyme (i.e. its nuclear head). We also see that in the case of vowel-initial words, the first unit in the structure is an empty onset, i.e. one devoid of any melodic content. As advocated by the theory, all consonant-final words must contain an empty nucleus at the end of the domain in accordance with the Onset Licensing Principle (Kaye, 1990).[3] Let us now turn to a more complex type of phonological arrangement in a word which contains a consonant cluster. This is illustrated in Figure 2 below.

Figure 2. The structures of *ikrel* 'hold' and *kirvi* 'godmother'



The structure presented in Figure 2a contains the cluster [kr] which is represented as a branching onset domain where the consonants are attached to positions involved in a governing relation with the left-hand slot being the governor and the right-hand one its governee. Such a structural configuration

[3] The Onset Licensing Principle requires that each onset within a phonological domain must be governed by the following nucleus.

imposes certain limitations on the choice of potential governees. Specifically, segments attached to governed positions have to be melodically less complex than their governors. Melodic complexity, in turn, is calculated in terms of the number of elements segments are composed of. More precisely, in the word *ikrel*, [k]—the governor—contains the elements of stopness, noise, place and the laryngeal prime **H** (defining voicelessness), i.e. four elements. [r], on the other hand, is defined by only one prime **A** which specifies its coronal place of articulation. Thus, the required complexity slope is observed here. As a result, [kr] constitutes a legitimate branching onset cluster.

When we turn to the structure depicted in Figure 2b, the cluster [rv] is represented as a rhymal complement-onset domain which is also a governing domain. Yet, in this case, the direction of government is from right to left. Since, [v] occupies the governor's position, we expect it to be more complex than the segment associated to the governed slot, i.e. [r]. This condition appears to be satisfied as [v] is defined by the elements of place, noise, and laryngeal.

To summarise this section, let us briefly enumerate the most significant assumptions concerning the syllabification of consonant sequences. Thus, phonological constituents, i.e. onset, nucleus and rhyme, can be maximally binary branching. Any sequence of three consonants has to be either split by a nuclear slot (which is then empty, e.g. [trf]) or has to be represented as a domain composed of a rhymal complement followed by a branching onset structure (e.g. [ndr]). Word-internal clusters can be represented either as branching onsets or rhymal complement-onset domains, the choice depending on the complexity of the consonants involved. In any structural arrangement, the potential governor has to be no less complex than its potential governee in line with the Complexity Condition (Harris, 1994, p. 170). Words beginning with vowel sounds are represented as structures where the initial onset dominates an empty skeletal slot, which follows from the universal character of the Onset Licensing Principle. Further, in consonant-final items, the onset dominating this consonant has to be always followed by an empty nuclear position. Again, the same universal principle enforces such a structural setting.

In what follows, we shall closely examine the structure of North West Lovari Romani words listed in the RMS Database. We will try to determine consonant sequences allowed in two basic word positions—initial and final—with a view to discovering the most important principles of word structure adopted by the phonological system of Lovari. Subsequently, Lovari-specific parameters regulating the language's phonotactics will be put forward. Our

findings will be based on the examination of 953 sample phrases contained in the database which will be confronted with Peter Wagner's description of the language presented in *A Grammar of North West Lovari Romani* (2012).

5. THE WORD-INITIAL CONTEXT

5.1 THE DATA

In Lovari Romani, words can be initiated by both vowels and consonants, as substantiated by such words as *amal* 'friend' and *dad* 'father' respectively. In the former case, as already indicated, phonological domains will begin with an empty onset position. The inventory of consonants occurring word-initially includes: [l, r, n, t, d, k, g, p, b, f, v, t͡s, j, ʃ, J, s, z, t͡ʃ, χ, h, t^h, p^h, k^h]. This set might be supplemented by an additional sound [ʔ] which can also be attested at the beginning of words. Wagner (2012, p. 28) speaks of "its restriction to only initial position in a speech stream, independent of word boundaries, be it phrases, clauses or sentences, in absence of other consonants. In this position, vowels are realised with a preceding [ʔ], e.g., *aď vin* /aď in/ > [ʔa'ʝin] 'honey', *efta* /ejfta/ > ['ʔeɟ̩fta] 'seven'." Thus, apparently, Lovari Romani disfavours initial empty onsets and fills them in with a glottal stop. At this point we need to explain that within the framework of Government Phonology, empty skeletal positions, in order to remain silent, have to be suppressed by proper government coming from the following vocalic unit. A skeletal slot which fails to be properly governed, can be filled with a melody. This seems to be the case in Lovari vowel-initial words. Since the language allows word-internal consonant clusters, often pretty heavy, which are likely to be separated by empty nuclei in the lexical representation, we can hypothesise that Lovari Romani treats empty onsets differently than empty nuclei. The former seem to resist proper government, whereas the latter must be sanctioned by this type of relation. At this stage, however, such a statement can only be considered a stipulation as this problem requires a closer investigation into the licensing properties of nuclei in the language, which exceeds the bounds of the present analysis. In what follows, our attention will be focused on the attested consonant clusters that are allowed in the word-initial position. Table 1 specifies the legitimate cluster types together with the relevant examples.

Table 1. Lovari Romani word-initial consonant clusters

Cluster	Example	Gloss
pr	pre	on, for
p^{h}r	me phrāvav	I open
tr	trava	grass
kr	krabici	boxes
gr	grastes	horse
br	brada	beard
zr	zrovno (ta zrovno brršind delas)	it started to rain
zd	zdel	it seems (to me)
trh	trhos	market
tm	ande tma	in the dark
prst	prstos/o angušt	finger
p^{h}rl	phrla (phrla, s' oda kerrdjal?)	Oh brother!
krrl	krrlo	throat
krrstj	hi krrstjindo	he is baptised
brrš	brršind	Rain
str	stromos	tree
skl	skladinel	store
stj	stjena	wall
svj	svjetlos	light
šv	švedsko	Sweden
št	štar	four
šp	špiglos	mirror
sk	skamind	table
šk	škola	school
st	stolička	chair
sl	slušno	decent
hr	hruška	pear tree
hl	te hlidinel	guard
dj	o djiv	wheat, rye
dl	dluhi	debts
kl	kleji	keys
pl	ande zumin plavinenas o muxi	there were flies in the soup
nj	njišt	nothing
dv	dvora	yard
kv	kvuli	because (of)
rr	me rrovav	I cry
xv	xvalinel	praise
nm	te nmukhle	leave (sth)
mr	mro	My
čtvrtk	čtvrtkone muši	on Thursday we must
trrh	trrhos	market
vl	vlastno	own
vlk	vlken	wolves
zl	zlato	golden
zvl	zvlašno	strange
krm	krminas	feeding

krčm	krčma	pub
krrl	krrleha	throat
vžd	vždycki	always/every time

Note. Own elaboration.

The examples listed in Table 1 reveal that the range of the attested consonant sequences is pretty extensive. More specifically, the number of segments ranges from one up to six. In the case of sequences consisting of two consonants, we can find clusters representing the following types:

(6) *CC cluster types*

a. obstruent + liquid	pr, tr, kr, tl, pl, kl, br, gr, zl, vl, dl, hl, hr, sl, zr
b. obstruent + nasal	tm
c. obstruent + obstruent	xv, kv, dv, st, sp, sk, šk, šp, št, šv
d. nasal + nasal	nm
e. nasal + liquid	mr
f. liquid + liquid	rr

Strings of three consonants can consist of

(7) *CCC clusters*

a. obstruent + Obstruent + Obstruent	vžd,
b. obstruent + obstruent + liquid	zvl, str, skl, stj
c. obstruent + liquid + obstruent	trh, vlk
c. obstruent + liquid + liquid	p^hrl
d. obstruent + liquid + nasal	krm

Proceeding to four-member clusters, the following types can be attested:

(8) *CCCC clusters*

a. Obstruent + liquid + obstruent + obstruent	prst
b. Obstruent + liquid + liquid + liquid	krrl
c. Obstruent + liquid + liquid + obstruent	brrš, trrh
d. Obstruent + liquid + obstruent + nasal	krčm

As far as six member sequences are concerned, we can find a possible configuration of the [krrstj] type. It is not clear, however, whether the final glide symbol is in fact a separate segment or just the indication of the palatalised character of the preceding plosive. Still, we do have a cluster such as [čtvrtk] with three obstruents initially and two in the final position. In the forthcoming

discussion in 5.2, we shall try to interpret the evidence presented above from the perspective of Government Phonology.

5.2 THE ANALYSIS

When we look at the above data through the optic of the model of GP, the most significant question that we are confronted with is whether the consonant sequences which surface phonetically as clusters are indeed immediate neighbours in the phonological structure. Recall that the theory assumes that in the word-initial context, a phonological domain begins with an obligatory onset which can be either empty (in vowel-initial words) or melodically realised (in consonant-initial words). Such an onset constituent can contain maximally two consonants but the selection of onset members is not random. More specifically, a universal complexity slope between the left-hand and the right-hand consonant has to be observed. Segmental complexity, in turn, is calculable in terms of the number of elements a melody is composed of. The elements used in the representation of consonants include: stopness/occlusion, noise, place and the laryngeal prime, which specifies the type of phonation of the consonant. This can be either **L** (slack vocal cords/voicing) or **H** (stiff vocal cords/voicelessness). Additionally, onset segments have to satisfy the condition of the potential governee being defined by maximally two elements (see Harris, 1994). These restrictions seriously delimit the number of possible onset clusters. Thus, [kr, fl, tj] will be well-formed onset clusters, whereas sequences such as [km, ps] or [kt] will not. Below in (9) we summarise the major onset structure conditions concerning consonant cohabitation therein.

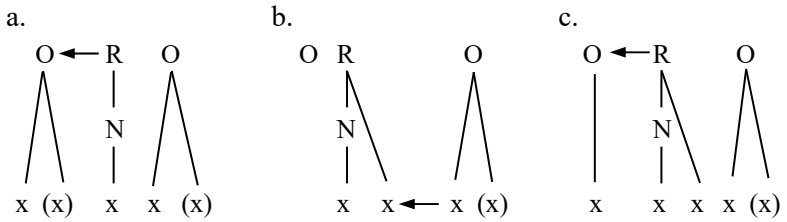
(9) *Onset structure conditions*

a. Onset structures are maximally binary branching
b. The left-hand member is an obstruent
c. The right-hand member is a liquid or glide
d. The onset positions are involved in a left-headed governing relation
e. The governor has to be more complex than the governee, the latter having maximally two primes
f. No place element sharing between the governor and the governee is allowed.

The restriction formulated in (9f) excludes clusters [tl] and [dl] from the set of possible onset strings.[4]

In the light of the discussion presented above, we see that Lovari Romani accepts branching onset structures. More precisely, out of the clusters listed in (6), the following sequences can be represented as branching onset domains: [pr, tr, kr, pl, kl, br, gr, zl, vl, sl, zr]. Thus, to the absence of alternations which would reveal that such consonant sequences are in fact separated by an empty nuclear slot, we will represent them as branching onset governing domains. Heavier consonant clusters can manifest the following structural configurations.

Figure 3. The syllabification of consonant clusters



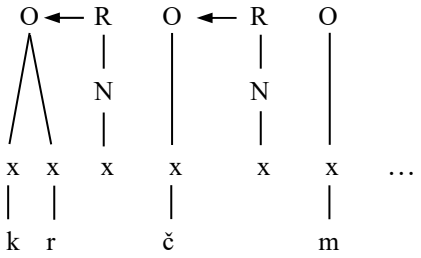
The structure in Figure 3a illustrates two optionally branching onsets separated by an empty nuclear position. When the onsets do not branch, we are dealing with a CC cluster whose members may not form a single intra-constituent governing domain because one or more of the onset well-formedness conditions (as defined in (9)) have been violated. In Lovari Romani, such a situation arises in words which begin with [tl, dl, hl, hr, tm, tn, xv, kv, dv, nm, mr, and rr]. Another alternative suggested in Figure 3a pertains to the situation when in a four-consonant cluster, we have two pairs of acceptable branching onsets. A closer look at the set of word-initial CCCC clusters listed in (6) reveals that we do not attest such structures in Lovari Romani. However, two additional structural settings are possible according to Figure 3a. Namely, three word-initial consonants can be represented as either a branching onset followed by an empty nucleus and another onset consonant, or a simplex onset followed by an empty nucleus and a branching onset structure. The former configuration is found in clusters such as [trh, vlk, p^hrl, krm], whereas no examples of the latter have been found in the Romani database.

[4] For arguments in favour of this claim, see Harris (1994).

Turning now to the structure illustrated in Figure 3b, it needs to be clarified that it depicts a very special type of relation called *magic licensing* by Kaye (1992) and attributed to *s*+C clusters. More specifically, a word-initial [s] is attached to the rhymal complement position but the nuclear head of this rhyme is empty. The fricative is governed by an onset position occurring to its right. The onset governor can be either simplex or branching. Thus, we will assign sequences such as [st, sp, sk, šk, šp, št, šv as well as str, skl, stj] to this group. The representation depicted in Figure 3c is an extended "variant" of Figure 3b in that the *s*+C sequence is preceded by some consonantal segment. When we examine the list of possible Romani word-initial clusters, the sequence [prst] seems to conform to this pattern though the onset preceding the domain of magic licensing is a branching one.

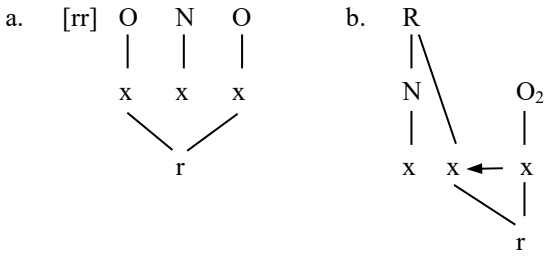
The cases discussed above cover the majority of the cluster types attested in Lovari Romani. Yet, within the set provided in (6), three heavy clusters deserve special attention. These are [krčm], [krrstj] and the heaviest [čtvrtk]. Let us analyse them one by one. In the case of [krčm], the two initial consonants constitute a well-formed onset, whereas [č] and [m] cannot belong to the same onset constituent since the nasal is too complex to be an onset governee. Hence, these consonant segments have to be dominated by separate onset positions. The onsets, in turn, need to be separated by an empty nucleus. We predict that the structure of the [krčm] cluster is as follows.

Figure 4. The structure of [krčm]



In [krrstj], two issues appear to be problematic. Firstly, the cluster seems to contain a consonantal geminate [rr] which, within the current framework, can be represented in one of two possible ways. The geminate can be either a sequence of two onsets separated by an empty nucleus or a single melody attached to the rhymal complement and the following onset points. The two options have been depicted in Figure 5a and Figure 5b respectively.

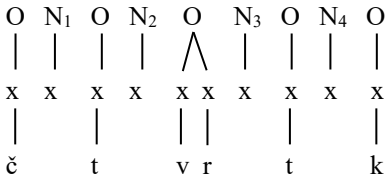
Figure 5. Geminate structures



At this stage, it would be difficult to decide which of the two structures is employed by the system of Lovari Romani. This problem calls for a much more detailed investigation into the behaviour of consonantal segments in this language. The data available in the Romani database are definitely insufficient. Hence, we leave this question open. The other problem concerning the cluster in question pertains to the status of the glide [j] as it is used in the database examples. It is not quite clear whether the symbol means a separate consonantal segment or serves to indicate a palatalised nature of the consonant it follows. On the former interpretation, [tj] would constitute a well-formed branching onset structure, while on the latter, just a simplex onset dominating a palatalised alveolar plosive. On either interpretation, however, the [stj] sequence can constitute a magic licensing domain with [s] attached to the rhymal complement position and the following consonant(s) being licensed by the onset to its right. This situation is depicted in Figure 3b.

The most complex structure seems to be involved in the representation of the cluster [čtvrtk]. Considered pairwise, only the [v] and [r] consonants can constitute a branching constituent domain. All the other segments have to be represented as separate onsets. This leads us to the following phonological representation of [čtvrtk].

Figure 6. The structure of [čtvrtk]



The problem this structure poses is a considerable number of empty nuclear positions since the nuclei N_{1-4} do not dominate any melodic content. The GP model assumes that empty positions have to be suppressed either through proper government or sanctioned parametrically in order to remain silent.[5] Since the cluster can be realised without any vowels between the consecutive consonants, we have to assume that the nuclei have been sanctioned in some way. Normally, in the case of proper government, it is a melodically expressed vowel that performs the function of a proper governor. Here, no such vowel is present. Hence, the only reasonable solution seems to be to propose that the onsets separated by empty nuclei form some sort of inter-onset governing domains which in many languages have the potential to suppress the realisation of the intervening empty nucleus.

To summarise our discussion of the word-initial context, it is noteworthy that Lovari Romani seems to employ a number of complex structural settings that phonetically yield heavy consonantal clusters word-initially. So in this context we can find single branching onsets and branching onset–simplex onset sequences separated by an empty nucleus. Lovari Romani also employs magic licensing domains underlying the *s*+C clusters as well as geminate structures. We can also attest combinations of the abovementioned types as observed in the case of the [krčm], [krrstj] and [čtvrtk] clusters. In the subsequent section, we shall examine the word-final context of Romani words to check whether similar structural configurations are also possible there.

6. THE WORD-FINAL CONTEXT

6.1 THE DATA

The examples supplied in the Romani database (RMS Database) clearly reveal that the language tolerates both vowels and consonants at the end of words. It is noteworthy, however, that no matter whether a lexical item phonetically terminates in a vowel or a consonant, its phonological representation will contain a nuclear constituent at the end of the domain. Recall that the GP model defines a phonological domain as a sequence of onsets and rhymes

[5] The Empty Category Principle requires that empty positions, in order to remain unrealised, have to be either parametrically licensed or properly governed. Proper government, in turn, is a stronger form of government. For details, see Harris (1994).

whose nuclear heads universally govern preceding onsets. In languages that have consonant-final words, the final nucleus will be empty as it is parametrically licensed. The *final empty nucleus parameter* is active in the majority of natural languages, e.g. Polish, French, English, German or Irish. As substantiated by the Romani data, it is also switched ON in Lovari Romani. Note the existence of such words as *vojakos* 'soldier' versus *baba* 'grandmother'. In the former, the phonetically final consonant [s] is in fact phonologically followed by an empty nucleus. Word-final empty nuclei have different licensing potential in different languages. The stronger they are as licensers, the more structure types they can support to their left. In order to determine the licensing strength of word-final empty nuclei in Lovari Romani, let us see which cluster types can be found before them. Table 2 summarises the available consonant clusters occurring at the end of Romani words.

Table 2. Consonant clusters in the word-final context

Cluster	**Example**	**Gloss**
rš	murš	man
rrš	berrš	year
ls	pridžarals	knew
Nj	e menj	neck
št	njišt	nothing
st	vast	hand, arm
ng	čhang	thigh
ps	e aps	teardrop
nd	brršind	rain
ndž	pandž	five
sk	lesk	him
štj	naštj	particle expressing inability

Note. Own elaboration.

As demonstrated in the table, the maximal number of consonants of a word-final cluster is two. Segment sequences attested in this context represent the following types.

(10) *Types of CC clusters*

a. liquid + obstruent rš, rrš (?), ls
b. nasal + obstruent ng, nd, ndž
c. obstruent + obstruent st, ps, sk, štj

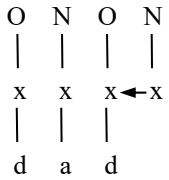
We have assigned the [štj] cluster to the (18c) category despite the fact that in the spelling three consonant letters have been used. The Romani database con-

tains only one example in which the cluster terminates the domain of the word. This suggests that the use of the [j] after the obstruent is merely a spelling convention that indicates the palatalised status of the consonant it follows. Similarly, the status of the [rrš] cluster is uncertain since the double [rr] is used only once in this word. The other occurrences of *berrš* contain a single [r] only.

6.2 THE ANALYSIS

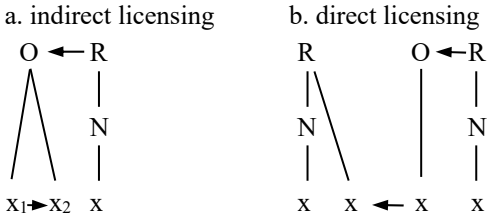
The evidence supplied in the Romani Database reveals that the domains of the language's words can terminate both in vowels and consonants, for example in *baba* 'grandmother' versus *dad* 'father'. In the former, the domain-final nucleus is naturally filled with a vocalic melody, whereas in the latter, the nuclear position remains empty. This indicates that the final empty nucleus parameter is switched on in the language. Thus, Romani licenses final empty nuclei which are allowed to remain unrealised. They are also strong enough to license the preceding onset material. Consider the structure in Figure 7 below.

Figure 7. The representation of *dad* 'father'



Moreover, not only onsets dominating single consonants receive such a licence from the right-hand empty nucleus but also more complex constituent structures. More specifically, as Lovari Romani tolerates two-segment clusters word-finally, it can be inferred that final empty nuclei entertain the status of *direct government-licensers*. It has to be explained at this stage that classical Government Phonology distinguishes between direct and indirect government licensers. A government licenser is a nucleus that authorises the preceding onset head to govern both the available onset complement position within a branching onset structure and the preceding rhymal complement position in the case of trans-constituent government. The two cases are illustrated in Figure 8.

Figure 8. Indirect and direct licensing paths



As indicated in Figure 8a, the nucleus, which is the head of the rhyme, licenses the preceding onset head x_1 to govern its complement x_2. This type of licensing is indirect. On the other hand, in Figure 8b, the onset head, which receives its government-licensing power from the following nucleus, governs the preceding rhymal complement position directly.

The Romani data provided in Table 2 indicate that, with a single exception of the [ps] cluster, word-final consonant sequences represent the Figure 8b type. Put differently, the preferred cluster type in this context is one consisting of the liquid or nasal followed by an obstruent. Such clusters are represented as rhymal complement-onset governing domains. This kind of government requires a government-licence from the following nucleus which in consonant-final words is melodically empty. Thus, it should be concluded that in Lovari Romani, word-final empty nuclei are direct government-licensers. Since branching onsets seem to be absent from the final context, we propose that final empty nuclei may not government-license indirectly.

The Romani Database contains a single example of a word terminating in the [ps] cluster (i.e. *e aps* 'teardrop'). It is noteworthy that such a sequence cannot be represented as a branching onset domain since [s], a potential onset complement, is an obstruent and obstruents are not allowed to be onset governees. The cluster may not manifest a rhymal complement-onset domain either as the fricative is less complex (in terms of the number of elements) than [p]. It seems, therefore, that in the [ps] cluster [s] should be analysed as the so-called appendix, or an extra-syllabic margin (Harris, 1994). In such a structural configuration, the fricative [s] is integrated into the phonological hierarchy at word-level, bypassing the level of constituents. In other words, it will not be syllabified into any of the constituents.[6]

[6] For an extended discussion of the concept of appendix, see Fudge (1969), Halle and Vergnaud (1980), or Kiparsky (1979).

CONCLUSION: A TYPOLOGICAL PERSPECTIVE

In this section we would like to compare the phonotactic patterns attested in the word-initial and final contexts in North West Lovari Romani words with those found in some other languages such as English, Polish and Irish. These systems have been chosen for comparison because they represent three different language families.

As for the left-hand edge of the word, Romani exhibits similar structural characteristics as the three abovementioned systems. More specifically, in all of the languages, words can start with empty onsets as well as simplex and branching onset constituents. Table 3 contains examples substantiating this assertion.

Table 3. Word-initial patterns in Romani, Polish, English and Irish

Pattern	**Lovari Romani**	**Polish**	**English**	**Irish**
Empty onset	ovca ‘sheep’	owca ‘sheep’	oak	ocras ‘hunger’
Non-branching onset	dad ‘father’	tata ‘dad’	dad	bolg ‘stomach’
Branching onset	pre ‘on’	prawy ‘right’	pray	bris ‘break’
Magic licensing (*s*+C)	stromos ‘tree’	strach ‘fear’	street	scoil ‘school’

Note. Own elaboration.

We can also see that the systems in question allow the so-called magic licensing domains word-initially. Thus, in terms of the permitted intra- and trans-constituent governing domains, the four languages seem to exhibit the same characteristics. Differences arise once we consider the length of word-initial consonant clusters. As mentioned above, Lovari Romani tolerates up to six consonantal segments at the beginning of words, though the corpus used in our study contains only one word with so many consonants in the word-initial context. The longest clusters frequently found there contain four consonants. In this respect, the language resembles Polish where the longest sequence comprises four segments, i.e.[vzgl, fskr, vzdr, vzbr, fstr] (Śledziński, 2010, p. 76). In English, we can have three consonants at the beginning of words, the leftmost of which has to be [s], as in *spray*, *splash* or *screen*.[7] Irish imposes the same restriction on cluster size as English. As mentioned in Ó Siadhail (1989, p. 20), the initial cluster may consist of maximally three

[7] The analysis of English consonant clusters within the domain of the word is presented in Harris (1994).

consonants, these being "a sibilant, plosive and a continuant sonorant".[8] Thus, in Irish and English, the longest word-initial clusters involve a domain of magic licensing. The situation in North West Lovari Romani and Polish is more complex. We predict that similar to Polish, the phonological representation of four member (and longer) consonant clusters should involve governing relations between onsets which would sanction/suppress intervening empty nuclei.[9]

When turning to the word-final context, certain significant differences between the four languages can be observed. The available structural settings for the attested consonant clusters occurring at the end of words have been summarised and illustrated in the Table 4.

Table 4. Word-final patterns in Romani, Polish, English and Irish

Pattern	**Lovari Romani**	**Polish**	**English**	**Irish**
Non-branching onset	e jag 'fire'	kot 'cat'	cat	gan 'without'
Branching onset	NO	wiatr 'wind'	NO	NO
Rhymal complement-onset domain	vast 'arm'	marsz 'march'	lamp	ard 'high'
Appendix	e aps 'teardrop'	klaps	eclipse	NO

As expected, all the four systems allow non-branching onsets at the end of words. Also, rhymal complement-onset domains are allowed in this context. These organise consonant sequences of the falling sonority profile, i.e. liquid/nasal/fricative + fricative/affricate/plosive strings. The condition that has to be satisfied here is a slope in elemental complexity between the potential governee (the left-hand segment) and its right-hand governor. More precisely, the potential governor must be no less complex than its governee. The languages differ in terms of the occurrence of branching onsets and appendices at the right edge of the domain. Specifically, out of the four systems, only Polish tolerates branching onset structures at the end of words. This is so because the type of governing relation binding the potential onset members requires indirect government-licensing from the following empty nucleus. Apparently, Lovari Romani, English and Irish empty nuclei are too weak to grant that kind of licence to the preceding onset head. With respect to the possibility of a word terminating in an appendix, Romani behaves like Polish and English in allowing such a structural solution, whereas Irish rules it out.

[8] See Bloch-Rozmej (1998) and Cyran (1997) for government-based analyses of word structure in Connemara Irish and Munster Irish respectively.

[9] For a detailed analysis of the consonantal system of Polish, see Gussmann (2007).

The examination of consonant clusters attested at domain edges in North West Lovari Romani has revealed certain interesting facts about the phonology of this system. We have determined the types of constituent structures employed by the language to organise consonantal material in the phonological representation of words. In particular, Romani uses non-branching and branching onsets at the beginning of words, with the latter being disallowed in the word-final context. However, rhymal complement-onset domains are found at both domain edges because they are easier to license by empty nuclei. Heavy consonant clusters, with three and more members, might require the recognition of a more stringent form of relation—inter-onset government—to represent them phonologically. This derives from the fact that onsets can be maximally binary branching, whereas any inter-consonant governing relation has to comply with the requirements of the Complexity Condition and that often rules out establishing either intra- or trans-constituent governing relations. As a result, we have to postulate empty nuclear positions to separate the phonologically illegitimate clusters. Since any empty nucleus, in order to remain phonetically uninterpreted, has to be suppressed either by proper government or inter-onset government, we would need to propose the existence of the latter to account for the attested forms. To make any definite judgments in this respect, a thorough investigation into the behaviour of Romani nuclei would have to be executed. This, however, goes beyond the scope of the present study.

REFERENCES

Bloch-Rozmej, A. (1998). *Element interactions in phonology. A study in Connemara Irish*. Wydawnictwo KUL.

Bloch-Rozmej, A. (2008). *Melody in government phonology*. Wydawnictwo KUL.

Boretzky, N. (2003). *Die Vlach-Dialekte des Romani. Strukturen – Sprachgeschichte –Verwandtschaftsverhaltnisse – Dialektkarten* [Romani Vlax Dialects. Structures – History of Language – Relationships – Dialect Maps]. Harrassowitz Verlag.

Charette, M. (1990). Licence to govern. *Phonology*, *7*(2), 233–253.

Charette, M. (1991). *Conditions on phonological government*. Cambridge University Press.

Cyran, E. (1997). *Resonance elements in phonology. A study in Munster Irish*. *PASE Studies and Monographs* (Vol. 3). Wydawnictwo Folium.

Cyran, E. (2003). *Complexity scales and licensing strength in phonology*. Wydawnictwo KUL.

Fudge, E. (1969). Syllables. *Journal of Linguistics*, *5*(2), 253–286.

Gussmann, E. (2007). *The phonology of Polish*. Oxford University Press.

Halle, M., & Vergnaud, J.-R. (1980). Three-dimensional phonology. *Journal of Linguistics Research*, *1*(1), 83–106.

Harris, J. (1994). *English sound structure*. Blackwell.

Kaye, J. (1990). 'Coda' licensing. *Phonology*, *7*(1), 301–330.

Kaye, J. (1992). Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics*, *2*, 293–313.

Kaye, J., Lowenstamm, J., & Vergnaud, J.-R. (1985). The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook*, *2*, 305–328.

Kaye, J., Lowenstamm, J., & Vergnaud, J.-R. (1990). Constituent structure and government in phonology. *Phonology*, *7*, 193–231.

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, *10*(3), 421–441.

Lakatošová, M., & Šebková, H. (2004). *Stručná mluvnice olašské romštiny* [A short grammar of Lovari Romani]. [Unfinished manuscript].

Matras, Y. (2002). *Romani: A linguistic introduction*. Cambridge University Press.

Ó Siadhail, M. (1989). *Modern Irish. Grammatical structure and dialectal variation*. Cambridge University Press.

Scheer, T. (2004). *A lateral theory of phonology: What is CVCV, and why should it be?* (Vol. 1). Mouton de Gruyter.

Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff & R. Oerhrle (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students*, 107–136. MIT Press.

Štatistický Úrad Slovenskey Republiky. (2001). *Demografia a sociálna štatistika.* April 11, 2023. Retrieved May 5, 2025, from https://slovak.statistics.sk/wps/portal/ext/themes/demography

Śledziński, D. (2010). Analiza struktury grup spółgłoskowych w nagłosie oraz w wygłosie wyrazów w języku polskim. *Kwartalnik Językoznawczy*, *3*(4), 61–83.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change.* Mouton.

Wagner, P. (2012). *A Grammar of North West Lovari Romani. Gramatika severozápadní olaštiny (lovárštiny)* [Doctoral dissertation, Univerzita Karlova v Praze].