Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

83

Tomislava Bošnjak Botica, Gordana Hržica

OVERABUNDANCE IN CROATIAN DUAL-CLASS VERBS

dr. sc. Tomislava Bošnjak Botica, Institut za hrvatski jezik i jezikoslovlje, tbosnjak@ihjj.hr, Zagreb
dr. sc. Gordana Hržica, Edukacijsko-rehabilitacijski fakultet, ghrzica@erf.hr, Zagreb

Croatian verbal inflection morphology is typically described using verb class distinctions. The number of classes differs among approaches, but the basic criterion for class division is the presence or absence and the type of suppletion in verb stems. Generally, one verb belongs to one inflectional class or paradigm only. However, some verbs belong to two classes, i.e. they have two parallel sets of stems. In such dual-class verbs, one infinitive form is realizable in two present forms in all cells within a class, i.e. there is an overabundance (Thorton 2011). Inevitably, one of the stem forming paradigms is a class with categorial suppletion. The present stem of a categorial suppletion class has a greater phonological distance from the infinitive stem than the present stem of the other class. Using a different terminology one class can be described as more transparent, while the other is less transparent (more opaque) in forming the present stem.

This study attempts to present overabundance in dual-class verbs and to determine whether competition in such forms can be explained by their tendency to conform to one default class or by other factors, specifically, by the phonological distance between the two paradigms of dual-class verbs. More than 90 dual-class verbs were retrieved from various sources, several dictionaries of the Croatian language and several corpora. Approximately half of the verbs were confirmed as dual-class verbs both in a dictionary and in at least one corpus. There are, however, verbs that were confirmed as dual-class verbs only in corpora – more precisely, in the web-based corpus hrWaC. Since they were attested as dual-class verbs only in corpora, they were treated as dual-class verbs that do not belong to the standard language base, but rather to sub-standard variants. They were therefore treated as a separate group in further analysis. While some dual-class verbs are more frequent in their non-suppletive class, others show preference towards the suppletive class, thus showing that the simplification of the paradigm towards the default class should not be considered the only guiding factor in the competition.

Two factors were examined as possible predictors of class preference: frequency and phonological distance. Overall verb frequency was calculated in the largest corpus available, hrWaC, using the lemma search. Phonological distance was calculated for each verb using the procedure based on Levenshtein distance (Levenshtein, 1965). The

84

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

distance is calculated on the basis of three operations: insertion, deletion and substitution, as described in the literature (overview and validation: Sanders and Chin, 2009). Only phonological distance was found to be a predicting factor of class preference.

These results point to two conclusions. First, they seem to speak in favour of the postulate of the theory of Natural Morphology, which states that unnatural language phenomena (suppletion included) occur more frequently in inflectional languages. Second, they can corroborate the claim (Dressler 1999) that strongly inflectional languages often have no default, or only a weak default, among competing morphological patterns. However, it is important to state that, within this theoretical framework, it has also been predicted that suppletion would be among the least transparent phenomena, and thus the most unnatural (weak suppletion being only one level lower on the scale than strong suppletion). This seems to suggest that verbs with greater phonological distance would more often appear in the non-suppletive class rather than the forms with less phonological distance.

Due to the relatively small amount of variance explained, additional factors will have to be considered in order to describe the variability in dual-class verb preference. Factors such as the frequency of the verb class, the number of verbs per class, frequency within a cell, and phonological conflicts arise as potential candidates. Also, more recent approaches to phonological distance use phonetic correlates instead of phonemes to determine the distance between two segments (e.g. Sanders and Chin 2009). Such fine-graded analysis might contribute to the explanation of this phenomenon.

Key words: *overabundance; competition; verbal inflection; suppletion; phonological distance; Croatian language*

1. Introduction

Several approaches have claimed that parallel forms with the same meaning are rarely (if ever) evidenced in languages. According to the blocking effect (Aronoff 1976), morphological doublets exclude each other. The principle of contrast (Clark 1987) claims that any two forms must contrast in meaning, and the Constant Rate Hypothesis (Kroch 1994) assumes some functional distinction[1] between similar forms. However, parallel forms with the same meaning are evidenced in languages: different forms of the same case in noun declension, different nominalization suffixes, parallel forms in verbal inflection, analytic and synthetic comparatives, etc. It has long been assumed that this phenomenon is rare in morphology, tending to be small in number and diachronically unstable (e.g. Kroch, 1994), but this view has been challenged in recent literature (e.g. Fehringer 2004, Thorton 2011).

---

[1] A sort of semantic or sociolinguistic distinction, e.g. figurative versus literal meaning, difference in register, social or dialect variants, etc. (Fehringer 2004).

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

85

1.1. Overabundance

Several terms (morphological variation/doubletism/synonymy/competition/ overabundance) have been used in literature to describe the phenomenon of parallel forms. In this study, the terms *overabundance* and *morphological competition* will be adopted to describe the Croatian verbal system. These two terms differ, and it should be understood that the overabundance that exists in some category will not necessarily result in morphological competition, nor will this competition necessarily result in the disappearance of one of the competing forms (Thornton 2012).

The notion of *overabundance* was introduced by A. Thornton (Thornton 2011, 2012) and stems from the approach of canonical typology introduced by Corbett (2005, 2007).[2] According to Corbett, a paradigm must have completeness, distinctiveness, predictability, and consistence to be canonical (Corbett 2007: 9). This means that, in order to have a canonical paradigm, it is expected that every single cell of the paradigm will be filled by the inflectional system (complete paradigm), each form will be different (distinctive cells), the stem will be predictable (regular), and the inflections will also be predictable (Corbett 2005: 33).[3] It is assumed that a single cell will be filled by a single inflected form (Thornton 2011: 360). Contrary to this, Thornton identifies paradigms in which there are at least two forms that share the same morphosyntatic properties in Italian verbal inflection (Thornton 2011: 359–360) and defines *overabundance*[4] as "the situation in which two (or more) inflectional forms are available to realize the same cell in an inflectional paradigm" (Thornton 2011: 358). These different forms in the same cell are called *cell-mates* (2011: 360), and they may differ in three ways: different composition/ structure; different lexical material; different inflectional endings (Thornton 2011: 361). The last of these has been well-confirmed in morphologically rich languages such as Croatian.

Overabundance has mostly been studied in Italian and some other Romance languages (Thornton 2011, 2012, Santilli 2014, 2016). But, if it is considered to be a part of the broader phenomenon of morphological dualism, then it has also recently been described in Dutch (Fehringer 2004), German (Mörth and Dressler 2014), Czech (Bermel and Knittl 2012, 2015), and Croatian (Lečić 2015).

1.2. Competition in overabundance

According to the models of morphological competition, it is inevitable for forms that occupy the same cell to compete, resulting in the decline of one of the forms.

---

[2] "We have looked at an approach to defining 'possible human language' and have given most attention to possible words" (Corbett 2005: 44).

[3] Along with canonical expectation, Corbett introduces a variety of real inflectional systems through corresponding morphological phenomena: defectiveness, syncretism, suppletion, and deponency (2005: 34).

[4] Coming from Ital. *sovrabbondante*, Lat. *abundantia*. Term used in descriptive grammars of Latin and Greek at least since the early 19th century to describe several kinds of nouns that present doublets in their paradigm (Thornton 2016).

Although the literature on overabundance does not predict the disappearance of one of the verbs, it does state that it is hardly ever the case that both members of the same cell display equal frequency in usage (Thorton 2012). In this sense, the two forms do compete, and one of them is more prevalent.

Theoretical models that can explain this manner of competition revolve around two approaches. One is polarized between one regular (default) paradigm and one irregular paradigm. This is based on the language processing approach, which aims to prove that frequent forms tend to be accessed directly rather than produced by a rule. According to this, irregular paradigms show evidence of overgeneralisation and paradigm change when frequency is taken into account (e.g. Pinker 1984). Consequently, frequent overabundant words will be more prone to using irregular patterns, while less frequent words will more likely conform to the regular (default) paradigm.

The other approach introduces language typology as a relevant factor, stating that, although morphological change is different in morphologically diverse languages, it is always governed by similar principles associated with the paradigm itself (such as morphotactic transparency (Dressler 2005) or morphological complexity (e.g. Dahl 2004)).

This study aims to present overabundance in a part of the Croatian verbal system and to provide insight into the factors that govern competition between the two overabundant forms.

1.3. Overabundance in Croatian

Different forms within one cell can be found in several morphological categories in Croatian. Thornton (2016) proposes four types of overabundance based on the number of overabundant lexemes and the number of cells that participate in the phenomenon. In some cases, overabundance appears in only one lexeme. For example, one overabundant lexeme would be *čovjek/ljudi* 'man/men', in which the overabundance is present in all the cells of the lexeme. Contrary to that, overabundance sometimes appears in a set of lexemes that is possibly defined by the inflection class or phonological or semantic properties. For example, there is a set of feminine nouns that can have two forms in dative/locative. The set of nouns is defined by phonological properties and the overabundance is present in two cells of the paradigm.

Overabundance in Croatian is evidenced in sets of lexemes in a different number of cells of a paradigm, or in the whole paradigm (examples in Table 1).

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

87

Table 1. Examples of overabundance in Croatian morphology

Overabundance of a set of lexemes in one cell	
NOUNS	**EXAMPLE**
Instrumental singular of masculine nouns[5] (a-declension)	*ribarom/ribarem* 'fisherman' *vijencom/vijencem* 'wreath'
Vocative singular of masculine and feminine nouns (a-declension, e-declension)	*Domagoj/Domagoju/Domagoje* *Patrik/Patriku/Patriče* *Sanja/Sanjo* *izdajico/izdajice* 'traitor'
Instrumental singular of feminine nouns ending in a consonant (i-declension)	*noću/noći* 'night' *ljubavlju/ljubavi* 'love'
Genitive plural of masculine nouns (a-declension)	*zubi/zuba/zubiju* 'teeth' *mravi/mrava* 'ants' *nokata/noktiju* 'fingernails'
Genitive plural of feminine nouns (e-declension)	*naranČā/naranČi/naranaČa* 'oranges' *nogu/noga* 'legs'
Genitive plural of neuter pluralia tantum nouns (a-declension)	*prsa/prsiju*[6] 'chest'

Overabundance of a set of lexemes in a set of cells		
NOUNS	**EXAMPLE**	
Nominative and Accusative singular of masculine nouns (a-declension)	*predio/predjel* 'district'	
Dative and Locative singular of feminine nouns (e-declension)	*toČki/toČci* 'dot', *rešetki/rešetci* 'grill'	
Plural of masculine gender (a-declension)	*golubi/golubovi* 'pigeons' *nosevi/nosovi* 'noses'[7]	
ADJECTIVES	**EXAMPLE**	
Genitive, Dative and Locative singular of masculine gender	*bratova/bratovog, bratovu/bratovom* 'brother's'	
PRONOUNS	**EXAMPLE**	
Genitive, Dative and Locative singular of masculine gender	*mojeg/mog, mom/mojem* 'mine' *njegova/njegovog, njegovu/njegovom* 'his'	

---

5   Presented in Lečić 2015.

6   The form *prsiju* is considered non-standard, although it is well documented both in literature and in corpora.

7   Of course, only forms not considered dialectal idiosyncratic variants are mentioned here. The quantity of the examples of the phenomenon would be significantly augmented were they included.

88

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

Overabundance of a set of lexemes in all the cells of the paradigm	
NOUNS	**EXAMPLE**
Set of nouns	*estet/esteta* 'aesthete'
ADJECTIVES	**EXAMPLE**
Comparatives	*dulji/duži* 'longer'
PRONOUNS	**EXAMPLE**
Possessive pronoun of the feminine gender	*njen/njezin* 'her'
VERBS	**EXAMPLE**
Present stem	*skitam/skićem* 'wander$_{pres.1st.sg}$' *znam/znadem* 'know$_{pres.1st.sg}$'

These examples illustrate overabundance in several Croatian subsystems, but do not exhaust this phenomenon. Overabundance is evidenced in derivational morphology as well, e.g. in nominalization, the formation of diminutives and augmentatives, and colour naming. A whole range of parallel forms also exists if accentuation is taken into account.

This study deals with overabundance in a part of the Croatian verbal system, i.e. in verbs with dual-class membership.

1.4. Suppletion in Croatian verb classes

Croatian verbal inflection morphology is typically described with the help of verb class distinctions.[8] The starting point for categorization is the difference between the two stems of a verb: the infinitive stem and the present tense stem.[9] Since the approach employed here will treat the thematic vowel (and more elaborate changes in verb forms) as a part of the stem, the basic criterion for class division will be the presence or absence of suppletion[10] in verb stems and its type. There are multiple approaches to suppletion (overview: Veselinova 2006); however, Croatian verbal classes display categorical suppletion according to tense, since in some verb classes one stem is used for non-present tense forms, while the other two are used for the present tense forms (Veselinova 2006).

---

[8] There are several classifications and the most commonly used are presented in Babić et al. 1991, Barić et al. 1997, Jelaska 2005, Silić and Pranjković 2005. For a detailed overview with the latest classification, see Bošnjak Botica 2011.

[9] Infinitive stem and present stem are used for more than just forming the infinitive and present, but the two will be used as examples when discussing verb class.

[10] On suppletion, see Meľčuk 1994, Hippisley et al. 2004, Corbett 2005, 2007, Veselinova 2006 etc. For his canonical approch, Corbett adopts Maľčuk's definition: For the signs X and Y to be suppletive, their semantic correlation should be maximally regular, while their formal correlation is maximally irregular (Meľčuk 1994: 358).

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

89

Table 2. Suppletion in Croatian verbal class*es*

		No suppletion	Categorial suppletion	
Infinitive		*hodati* hod-a-ti 'to walk'	*pisati* pis-a-ti 'to write'	*putovati* putov-a-ti 'to travel'
Present	1st singular	hod-a-m	piš-e-m	putuj-e-m
	2nd singular	hod-a-š	piš-e-š	putuj-e-š
	3rd singular	hod-a-ø	piš-e-ø	putuj-e- ø
	1st plural	hod-a-mo	piš-e-mo	putuj-e-mo
	2nd plural	hod-a-te	piš-e-te	putuj-e-te
	3rd plural	hod-a-(j)u	piš-ø-u	putuj-ø-u
Infinitive stem		hoda-	pisa-	putova-
Present stem		hoda-	piše- pišø-	putuje- putujø-

Suppletion has been viewed as a gradable phenomenon (Dressler 1985, Mel'čuk 1994). S*trong*, *genuine*, or *full* and *weak*, *quasi*, or *partial* (e.g. Dressler 1985, Mel'čuk 1994, Corbett 2005) can be seen as two points on the continuum. Strong suppletion is an alternation in which there is no shared phonological material between the suppletive forms, where a single lexical item has completely different stems (čovjek/ljudi 'man/men'). Weak suppletion assumes an alternation in which two forms share some phonological material, that is, only a part of the form changes (kleti/kunem 'to curse/curse$_{pres.1st.sg}$').[11] There is a continuum of lowering suppletivity, from strong to zero suppletion, with infinitive intermediate cases (Mel'čuk 1994: 379) that may show greater or smaller phonological distance. In the examples in the table, it is evident that the verb hodati 'to walk' shows no suppletion, while pisati 'to write' and putovati 'to travel' show categorial suppletion. However, there is a greater phonological difference in forming the two suppletive stems for the verb putovati 'to travel' than in forming the two suppletive stems of the verb pisati 'to write'.

Example 1. Phonological distance in two verbs

	Phonological distance	
	Pisati pis-a-ti 'to write'	*putovati* putov-a-ti 'to travel'
Infinitive stem	pis-**a**-	put**ov**-a-
Present stem	piš-**e**-	put**uj**-**e**-
Phonological changes	- thematic vowel - **one** phoneme	- thematic vowel - **two** phonemes

---

[11] As for terminology, some authors tend to speak about suppletion between phonologically dissimilar roots associated with the same lexical item, but grammaticaly or lexically conditioned allomorphy between phonologically dissimilar realisations of the same morphosyntatic property (Carstairs-McCarthy 1987).

90

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

Generally, each Croatian verb belongs to one inflectional class or paradigm, whether it is a class with no suppletion or a class with categorial suppletion. However, some verbs belong to two classes, i.e. there are two parallel sets of stems forming the present tense. In such dual-class verbs, one infinitive form is realizable in two present forms in all cells within a class. Inevitably, one of the stem-forming paradigms is a class with categorial suppletion. The present stem of the categorial suppletion class has a greater phonological distance from the infinitive stem than the present stem of the other class. To use different terminology, one class can be described as more transparent while the other is less transparent (more opaque) in forming the present stem.[12] The phonological distance in this case reflects transparency/opaqueness because greater distance presents more challenges in the formation of the proper form of a verb.

Example 2. Two classes of the dual-class verb *šetati* 'to stroll'

		šetati šet-a-ti 'to stroll'	
		No suppletion	**Categorial suppletion**
Present	1st singular	šet-a-m	šeć-e-m
	2nd singular	šet-a-š	šeć-e-š
	3rd singular	šet -a-ø	šeć-e-ø
	1st plural	šet -a-mo	šeć-e-mo
	2nd plural	šet -a-te	šeć-e-te
	3rd plural	šet -a-(j)u	šeć-ø-u
Infinitive stem		šet-a-	šet-a-
Present stem		šet-a-	šeć-e
Phonological changes		- no change	- thematic vowel - one phoneme

The goal of this study is to identify and describe overabundance in the Croatian verbal system, to describe competition between the two paradigms in this kind of verb, and to determine whether this manner of competition can be explained by the tendency to conform to one default class or by other factors – specifically, the phonological distance of suppletive forms. It was predicted that the on-going competition in the class preference of dual-class verbs would not always result in the prevalence of a less complex form, and that there would be no frequency effect in class preference, showing that there is no tendency towards the default usage of one-stem class. However, phonological distance of verb class should predict class preference.

---

[12] Stump and Finkel (2013) describe class A as more transparent than class B if the number of principal parts required to deduce a given cell in A's paradigm is, on average, lower than the number of principal parts required to deduce a given cell in B's paradigm, if there are more alternative principal-part analyses for class A than for class B, and if A's paradigm has fewer unpredictable cells than B's paradigm.

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

91

2. Methodology

2.1. Materials

Two types of materials were used to identify dual-class words. Dictionaries of the Croatian language were used in order to restrict the findings to the standard language, which is the most widespread language variant. Apart from identifying individual verbs, the use of dictionaries also allows the detection of morphological subclasses in which dual-class verbs appear. Corpus based research included (1) searching for dual-class verbs already identified in dictionaries and (2) searching for other members of morphological subclasses that could be dual verb candidates.

2.1.1. Dictionaries

Several dictionaries published within the last twenty years were used for the extraction of dual-class verbs: the Large Croatian Language Dictionary (Anić and Jojić 2003) is a source with more than 190,000 contemporary Croatian lemmas; the Croatian Language Dictionary (Šonje and Nakić 2003) describes 64,000 lemmas; the School Dictionary of the Croatian Language (Birtić et al. 2012) consists of around 30,000 lemmas.

Two more dictionaries were used to identify dual-class verbs. Both of them have a historical component and do not represent the current state of the Croatian standard language. However, they contribute an interesting historical perspective and a large word base. The largest available dictionary of the Croatian language is the Dictionary of the Croatian or Serbian Language, which was published in 97 volumes (23 books) from 1888 to 1976. The dictionary consists of more than 400,000 lemmas, also including some words from Croatian non-Štokavian dialects and historical documents. The Croatian-Italian Dictionary (Parčić 1901, reprinted 1995) includes data from other dictionaries from that period, as well as words from the more current Croatian literature. These two dictionaries served to detect morphological verbal subclasses in which dual forms appear, but due to their developmental time span they could not be used as the only source to confirm the current standard dual-class usage of verbs.

2.1.2. Corpora

Four corpora of different origins, sizes, and functions were used to obtain a broad overview of the dual-class usage of verbs.

The largest source was the Croatian Web Corpus 2.0 (hrWaC). It consists of 1.9 billion tokens collected from the .hr top-level domain (Ljubešić and Erjavec 2011, Ljubešić and Klubička 2014). The trend of building web-based corpora has allowed the creation of large corpora for the so-called small languages. However, it must be taken into consideration that web corpus acquisition is much less controlled than that of traditional written language corpora, which call for better control in content analysis. HrWaC has been lemmatized and morphologically annotated. This corpus was selected

92

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

for this research because it is the largest corpus of the Croatian language currently available in terms of token size, in terms of different genres, and in terms of differences in the professional status of text writers.

Croatian National Corpus 3.0 (CNC) is a collection of selected texts written mainly in contemporary Croatian (Tadić 2009). At the moment, it has 216.8 million tokens. The majority of the texts are informative (74% from newspapers, magazines, journals, brochures, non-fiction books), 23% are fiction texts, and 3% are texts of mixed genres. This corpus was selected as a research source because it has the status of a national corpus and claims to represent the written foundation of the Croatian standard language.

The Croatian Language Repository (CLR) contains selected texts from various functional domains and genres of the Croatian language (Ćavar and Brozović Rončević 2012). Although it includes mostly written sources from the period of the beginning of the final shaping of the standardisation of Croatian (approximately the second half of the 19th century) up to more recent texts, it also hosts older texts. Texts are divided into two subcorpora: the literature subcorpus and the periodical subcorpus. The literature subcorpus was used in this research. At the moment, this subcorpus has approximately 20 million tokens. It consists of fundamental Croatian fictional texts, novels, dramas, and poetry from 1569 to 1946. This subcorpus was used as a source because it offers a wide diachronic perspective.

The Croatian Adult Spoken Language Corpus (CASLC) is a corpus of informal conversational spoken language (Kuvač Kraljević and Hržica 2015). It consists of 15-minute language samples of conversations between 2 to 6 participants recorded in different parts of Croatia. Currently, it consists of 300,000 tokens. It was used as a source because it is the only corpus offering spoken language data, as well as large dialectal diversity.

2.2. Procedure

Verbs that appeared in two different paradigms in any dictionary or in any corpora were selected and further treated as dual-class verbs. It was not necessary for the verb to appear in all morphological forms of a certain class: one occurrence in any form in both classes was the minimal criterion for establishing dual-class membership.

Example 3. Identification of dual-class verbs: evidence of morphological forms in the corpora

Infinitive	ŠETATI ('to stroll')		HIHOTATI ('to giggle')		PLAKATI ('to cry')	
	class 1	**class 2**	**class 1**	**class 2**	**class 1**	**class 2**
present 1st person singular	šetam	šećem	hihotam	hihoćem		plačem
present 2nd person singular						plačeš
present 3rd person singular	šeta	šeće				plače
present 1st person plural	šetamo					

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

93

Infinitive	ŠETATI ('to stroll')		HIHOTATI ('to giggle')		PLAKATI ('to cry')	
	class 1	**class 2**	**class 1**	**class 2**	**class 1**	**class 2**
present 2nd person plural		šećete				plačete
present 3rd person plural	šetaju	šeću	hihotaju	hihoću		plaču
Dictionary	+	+	-	+	-	+
Corpus	+	+	+	+	-	+
DUAL-CLASS VERB	+		+		-	

For each selected verb, several categories were outlined: frequency of appearance in the more transparent and the more opaque class, overall verb frequency, and the phonological distance between the two classes (to calculate how much the classes differ).

2.2.1. Class frequency

To obtain class frequency, each verb was searched for in every corpus. The sum of all frequencies of morphological forms belonging to a certain class represents the overall frequency of a certain class. Due to the fact that some of the corpora are not lemmatized, and to the fact that some dual forms are not standard, part of this analysis was done by searching for each morphological form separately. For others, the lemma search option with node form overview was used (NoSketchEngine). Eventually, each dual-class verb was associated with information about its frequency in both classes in all sources, together with the information on the percentage share of a certain class. This percentage determines the class preference of the verbs.

Example 4. Calculating class frequency and class preference

Infinitive	ŠETATI ('to stroll')			
	class 1 - morphological forms	frequency of the morphological form	**class 2** - morphological forms	frequency of the morphological form
present 1st person singular	šetam	87	šećem	48
present 2nd person singular				
present 3rd person singular	šeta	145	šeće	65
present 1st person plural	šetamo	54		
present 2nd person plural			šećete	4
present 3rd person plural	šetaju	132	šeću	42
class frequency		418		105
class percentage		80%		20%
class preference		+		

94

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

2.2.2. Verb frequency

Overall verb frequency was calculated in the largest corpus available, hrWaC, using the lemma search. Due to the fact that some dual forms are not standard, or were not recognised as belonging to a certain lemma, part of this analysis was done by searching for each morphological form separately and adding the frequency of such forms to the lemma frequency to calculate the overall frequency.

2.2.3. Phonological distance

Phonological distance was calculated for each verb using the procedure based on Levenshtein distance. Levenshtein distance (Levenshtein 1965) compares the difference between two sequences, and due to its applicability to any sequence of symbols, it has been implemented in many fields, including linguistics. The distance is calculated on the basis of three operations: insertion, deletion, and substitution. Kessler (1995) and following works (overview and validation: Sanders and Chin 2009) treat insertion and deletion as simple operations (both of which are weighed 1 point), while substitution includes both the deletion and insertion of a new element, and is therefore considered to be a complex operation (weighing 2 points). This is not an approach pursued by all, but there seems to be significant evidence in favour of this kind of calculation (e.g. Sanders and Chin 2009 found a correlation of .97 between the algorithm and native speaker judgment of phonological distance). To calculate phonological distance, an automated Levenshtein distance calculator tool was used (Yarkoni, Balota & Yap 2008). Values were set to deletion = 1, insertion = 1, and substitution = 2. As longer words tend to be separated by more distance than shorter words (Nerbonne, Heeringa and Kleiweg 1999), the distance of each word pair was normalized by dividing it by the length of the longer word.

It is worth noting that the phonological distance of individual verbs reflects the distance between two classes of the same verb. The goal was to establish a criterion that shows how similar or how different the two competing class types are.

2.2.4. Predicting class preference

Overall verb frequency and class type distance were factors that were considered to be possible predictors of class preference. A simple linear regression was calculated to determine whether overall frequency and class type distance can predict class preference.

3. Results

3.1. Number of dual-class verbs

The research has identified 92 dual-class verbs in all the available corpora. Most of the dual-class verbs were found in the largest corpus, hrWaC (90 – 98%). The second-largest source was the Croatian Language Repository (literature corpus) with 50 dual-

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

95

class verbs (60%). A Similar number was found in the dictionaries of the Croatian language (51 – 55%). The lowest number of written attestations of dual-class verbs was found in the Croatian National Corpus (24 – 26%). Only a small number of verbs were confirmed as dual-class verbs in the Croatian spoken adult language corpus (5 – 5%).
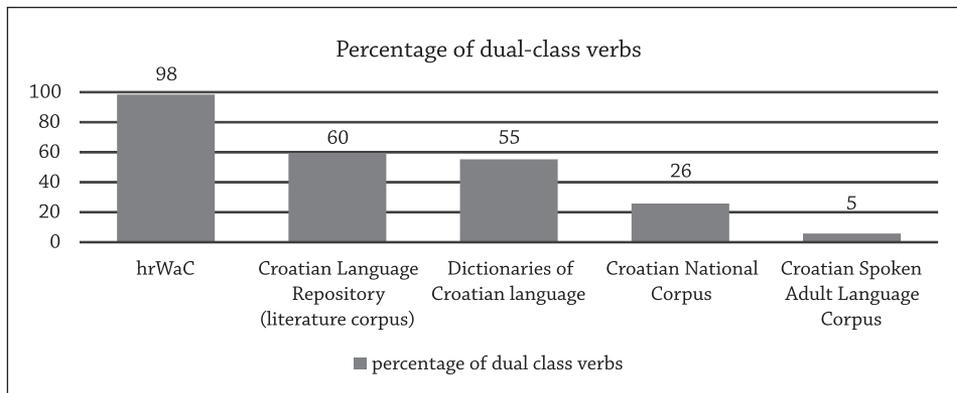
Figure 1. Percentage of all identified dual-class verbs per source

Half of the verbs were confirmed as dual-class verbs both in a dictionary and in at least one corpus. There are, however, verbs that were confirmed as dual-class verbs only in corpora, more precisely, in the web-based corpus hrWaC.

If at least two sources are considered relevant criteria for determining the duality of a given verb, the largest number of verbs (50) was found both in dictionaries and hrWaC (40 of them were also found in the Croatian Language Repository), while 42 verbs were found both in hrWaC and the Croatian Language Repository (40 of them were also found in the dictionaries). As for the spoken language corpus, dual-class verbs also appeared in some other source in the majority of cases (4 out of 5).

Table 3. Number of dual-class verbs in multiple sources

Corpora	Confirmations																
hrWaC	+	+	+	+	-	-	-	+	+	+	-	-	-	+	+	+	+
CLR	+	+	+	-	+	+	+	-	-	-	-	-	-	+	+	+	-
Dictionaries	+	+	-	-	+	+	-	+	+	-	+	+	-	+	+	-	-
CNC	+	-	-	-	+	-	-	+	-	+	+	-	+	+	-	-	-
CSALC	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
Number of dual-class verbs	22	40	42	**90**	0	0	1	24	**50**	24	0	1	0	2	1	1	1

96

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

It is important to state that there are 40 verbs that were only confirmed as dual-class verbs in hrWaC, the web-based corpus. They were attested in the corpora but not in the dictionaries, and thus do not belong to the standard language base, but rather to sub-standard variants. They were therefore treated as a separate group in further analysis. Contrary to this, another group of verbs appeared both in the dictionaries and in the corpora. The first group of verbs will be referred to as corpora dual-class verbs, while the other group will be referred to as dictionary-corpora dual-class verbs.

Table 4. Two groups of dual-class verbs

	Dictionary-corpora dual-class verbs	Corpora dual-class verbs
Attested in dictionaries	+	-
Attested in corpora	+	+

3.2. Frequency of dual-class verbs

The frequency of dual-class verbs shows great variability: the minimum frequency is 5 and maximum is 2,816,152. The mean result is 86,692, but the median is 1,138 with a standard deviation of 397,320. However, the majority of these verbs (67) have frequencies lower than 5,000.
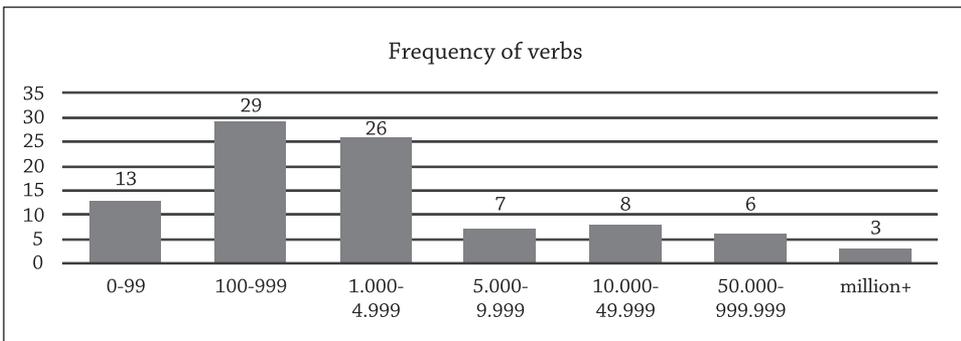
Figure 2. Frequency of dual-class verbs

When the two groups of verbs are compared, dictionary-corpora verbs show more prevalence in lower frequencies as compared to corpora verbs. Although the mean result is higher for dictionary-corpora verbs (127,854 vs. 35,489), this is the result of 3 outliers, and the median shows that there are more verbs with lower frequency in this group (1107 vs. 1168). However, there were no significant differences between the two groups (t(92) = -1.110, p = .27).

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

97
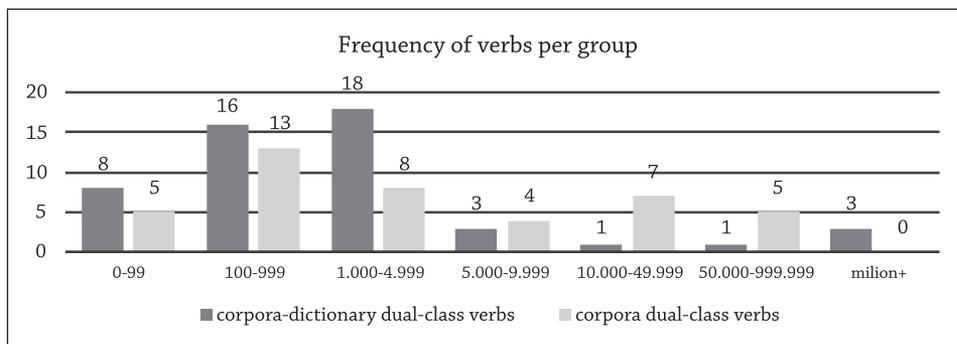
Figure 3. Frequency of dual-class verbs in dictionary-corpora dual-class verbs and in corpora dual class verbs

3.3. Class preference

The overall frequency of all types in a certain class allows the determination of class preference towards the more transparent or the more opaque class. For dictionary-corpora verbs, there is only a minor prevalence of the tendency to choose the more transparent class. The corpora verbs show a great preference for the more opaque class, i.e. they are generally verbs present in the standard language lexical base only in the class type with stems that exhibit great phonological distance, but they appear in corpora in a low percentage in the class type in which the stems are phonological equivalents.
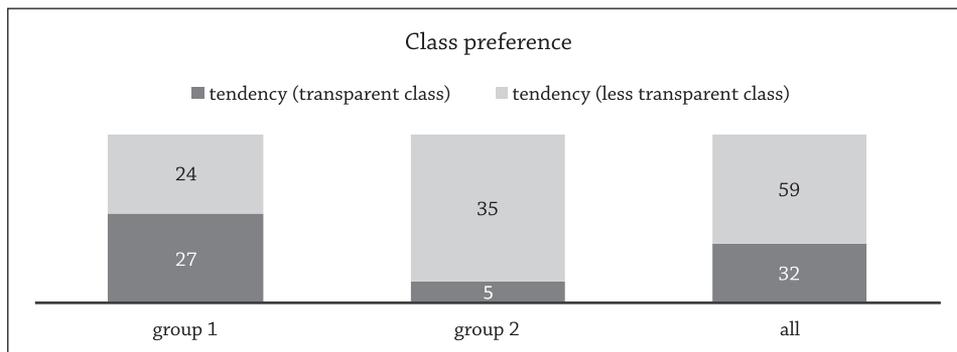
Figure 4. Class preference in dictionary-corpora dual-class verbs and in corpora dual-class verbs

A more precise measure of class preference is the percentage of types in one of the classes. When observing the percentage of types in the more transparent class, it is easy

to see that dictionary-corpora verbs show much more variability. While the majority of verbs tend to group around one of the two poles (extremely small percentage or extremely large percentage), there are still verbs along the continuum. This is not the case with corpora verbs: all but two verbs are on the final points of the continuum, and the majority of them (83%) have only 1-10% of their forms in the class with low phonological distance (or no distance at all). The two groups are significantly different (t(92) = -4.379, p < .05).
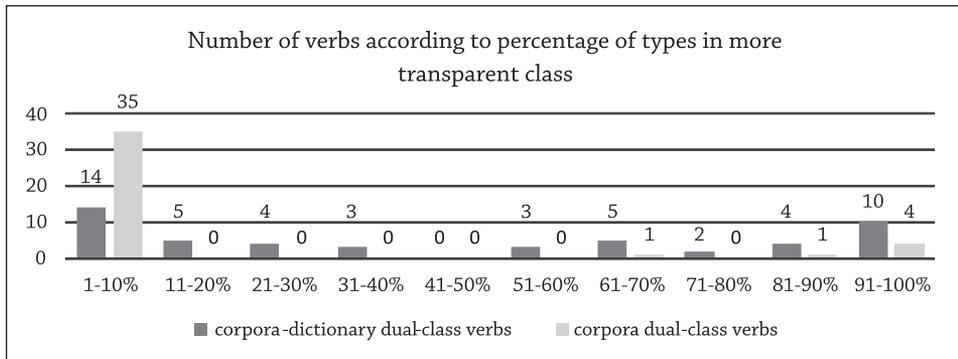
Figure 5. Number of verbs according to percentage of tokens in the more transparent class

It is important to state that the corpora verbs are considered non-standard dual-class verbs, but are attested in the dictionaries as one-class verbs. The frequency of forms in one of their classes (the one attested in the dictionaries) is, in general, extremely high (at least 90%), but the occurrence of forms in their other class shows that a new paradigm is beginning to appear. However, due to the low frequency of forms in the non-standard second class, which shows that competition has only started to emerge, they cannot contribute to class-preference research.

3.4. Phonological distance

All dual-class verbs have one class with greater phonological distance (more opaque) and one with lesser phonological distance (more transparent). Both corpora-dictionary dual-class verbs and corpora dual-class verbs most often have a phonological distance of 4 between classes, but corpora-dictionary class verbs are more evenly spread across other distance values. It is evident that the results of the raw measures of phonological distance partially reflect the morphological patterns in which verb stems are formed (verb classes).

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

99
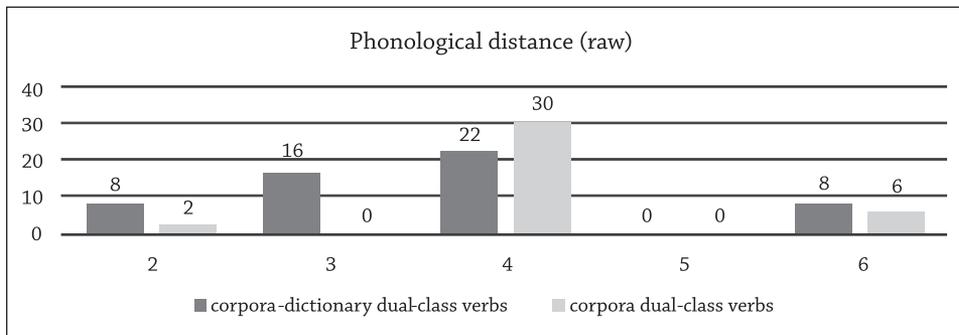
Figure 6. Number of verbs by phonological distance (raw)

The results of relative phonological distance show a wide range of results, but similar mean, minimum, and maximum values. The two groups do not differ (t(92) = -1.667, p = .099).

	Mean	Minimum	Maximum	SD
Corpora-dictionary dual-class verbs	54.69	17	120	20.03
Corpora dual-class verbs	61.51	25	120	18.86

3.5. Predicting class preference

Some dual-class verbs are more frequent in the more transparent class, while some are more frequent in the more opaque one. One factor that might explain class preference is the overall frequency of the verb (frequency effect should show that more frequent verbs will more often appear in the less transparent class while less frequent verbs will appear more often in the more transparent class). Another factor that might explain class preference is the distance in transparency (measured by phonological distance) between the two classes of the same verb.

Hierarchical linear regression was calculated to determine how predictive (1) frequency and (2) phonological distance were of class preference. Only verbs attested both in the dictionaries and in the corpora were put into the equation. They are dual-class verbs in standard Croatian, and it is therefore safe to claim that the duality of their forms is not the result of horizontal or vertical language variability reflected in the large corpora.

3.5.1. Predictiveness of frequency and phonological distance

Overall frequency was observed as a possible predicting factor for class preference. Overall frequency is the frequency of a verb in the hrWaC corpus. Class preference is shown by the percentage of usage in the more transparent class. The goal was to

100

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

determine whether frequency can predict the percentage of usage of a dual-class verb in its more transparent class.

The frequency of dual-class verbs and the relative phonological distance between the two classes of dual-class verbs were observed as possible predictors of class preference. In order to explain the role of multiple factors in class preference, a hierarchical regression was performed by adding the frequency as the first factor and relative phonological distance as the second factor.

The hierarchical multiple regression has revealed that, at stage one, frequency did not contribute significantly to the regression model ($F(1,90) = 1.866$, $p = .175$, $r^2 = .020$). Introducing relative phonological distance explained 18.7% of the variation ($F(2,89) = 8.924$, $p < .001$, $r^2 = .187$). Bias towards the more transparent class is higher if the relative phonological distance between the two classes of dual-class verbs is lower ($B = -.303$).

A lesser phonological distance between the two classes of dual-class verbs (i.e. more similarity between the two classes) predicts the more frequent usage of a verb in the transparent class, explaining 19% of the variance. A greater phonological distance between the two classes of dual-class verbs (i.e. less similarity between the two classes) predicts the less frequent usage of a verb in the transparent class.

4. Discussion

Due to the fact that dual-class verbs belong to specific morphological subclasses, the verbs found in this research do not exhaust this phenomenon. There are morphological subclasses that host many dual-class verbs found both in dictionaries and in corpora, but which also accommodate new corpora dual-class verbs (e.g. šetati – šetam – šećem 'to stroll' vs. šaputati – šaputam – šapućem 'to whisper'). New morphological subclasses that accommodate only corpora dual-class verbs have also been attested (e.g. pljuvati – pljuvam – pljujem 'to spit'). All of these subclasses might accommodate verbs that were not retrieved in this research.

Overabundance in dual-class verbs in Croatian shows a high level of variability. Dual-class verbs appear as a part of the standard variety, but also outside of it. They are present in different language sources and in different genres. The overall frequency of such verbs is shown to have a wide range, as is the frequency in one of the classes. Class preference differs among verbs: verbs do not always conform to the more or less transparent class. Almost half of the verbs attested as dual-class verbs in both the dictionary and the corpora prefer usage in the less transparent class. The group of verbs confirmed as dual-class verbs only in the corpora displays a greater tendency towards the class with more phonological distance between the two stems, but only because this is the regular class for the majority of them. Duality is presented in the low number of tokens as a sort of simplification of the paradigm. This is true for all morphological subclasses of corpora-only dual-class verbs that were identified in this research, except for verbs ending in -avati, which sometimes use a more complex paradigm (povećavati – povećavam – povećajem 'to enlarge'). Although this usage is restricted to certain dialects, it still

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

101

contributes to the description of this phenomenon, showing that a simplification in paradigm is not the only path to overabundance in the dual-class verbs[13]. Diachronic perspective might offer more insight into the nature of these verbs (Thorton 2012), but their duality cannot simply be seen as a journey towards a simplification in paradigm: the verbs that form the core of this research prefer both the more and less transparent classes.

Two approaches that offer predictions of morphological change have been taken into account in order to explain the class preference of dual-class verbs. Dual language processing models stem from the idea that very frequent forms tend to be accessed directly rather than produced by rule. According to this approach, irregular paradigms show evidence of overgeneralisation and paradigm change when frequency is taken into account (e.g. Pinker 1984). More frequent forms are accessed directly, and are therefore more prone to staying in the irregular paradigm. Less frequent forms, on the contrary, are more prone to conforming to the default (rule governed) class. The best candidate for this kind of class would be the verb class with no suppletion. The majority of dual-class verbs would be expected to conform to the default class, while more frequent dual-class verbs would conform to the irregular class (with more phonological distance between stems) and less frequent dual-class verbs would conform to the default class (with no phonological distance between stems). However, the results of this research do not confirm these claims, since the regression analysis has shown that frequency could not significantly predict class preference.

Other approaches, such as natural morphology, introduce language typology as a relevant factor, stating that suppletion and competition are different in morphologically diverse languages. While competition is generally viewed as a road towards a simpler paradigm, Dressler (1985) states that a language, depending on its type, will not necessarily choose the more natural option. In this sense, agglutinative languages are less prone to suppletion, while it appears quite often in inflectional morphology. The reason is the nature of these language types: agglutinative languages form long words with transparent rules, while inflectional languages prefer short and indexical words that are often morphotactically opaque. Following this, morphological forms in Croatian (which is a highly inflective language) should not always be prone to simplifications as a more natural option. Croatian dual-class verbs should not always choose the more transparent verb class, although general principles associated with the paradigm (e.g. morphotactic transparency – Dressler 1985) will guide the process of simplification when it occurs. When phonological distance, i.e. similarity between the two paradigms of a dual-class verb, was taken into account, it was a significant predictor of class preference. Smaller differences in the two paradigms of dual-class verbs predict a more frequent usage of a verb in the transparent class, explaining 19% of the variance. Greater differences in the two paradigms of dual-class verbs predict a less frequent usage of a verb in the transparent class.

---

[13]   As it is true, for example, in child language overregularisations (Hržica 2012).

102

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

5. Conclusion

This research into overabundance in Croatian has identified many dual-class verbs in several morphological subclasses. Some of these subclasses accommodate verbs attested only in corpora, indicating that the Croatian verbal system shows a persistent tendency towards new overabundant paradigms. This seems to suggests that overabundance (like suppletion, heteroclisis, etc.) is a genuine type of non-canonical phenomenon that can occur in paradigms, and that the idea occasionally posited in the literature that overabundance will eventually be inevitably eliminated in all cases is not fully supported by the Croatian data. Although there is a good candidate for a default class in Croatian (one-stem class), verbs do not always conform to this paradigm, and instead show a great deal of variability. Phonological distance, i.e. similarity between the two paradigms of dual-class verbs, partially explains class tendency.

These results point to two conclusions. First, they seem to speak in favour of the postulate of the theory of Natural Morphology, which states that unnatural language phenomena (suppletion included) occur more frequently in inflectional languages. Second, they can corroborate the claim (Dressler 1999) that strongly inflectional languages often have no default, or only a weak default, among competing morphological patterns. However, it is important to state that, within this theoretical framework, it has also been predicted that suppletion would be among the least transparent phenomena, and thus the most unnatural (weak suppletion being only one level lower on the scale than strong suppletion). This seems to suggest that verbs with greater phonological distance would more often appear in the non-suppletive class rather than forms with less phonological distance.

Due to the relatively small amount of variance explained, additional factors will have to be considered to describe variability in dual-class verb preference. Factors such as the frequency of the verb class, the number of verbs per class, frequency within a cell, and phonological conflicts arise as potential candidates. Also, more recent approaches to phonological distance use phonetic correlates instead of phonemes to determine the distance between two segments (e.g. Sanders and Chin 2009). A fine-graded analysis such as this might contribute to the explanation of this phenomenon.

This research showed only a part of overabundance in the Croatian verbal system. At least two more types of overabundance in verbal inflection are attested: lexemes that have two suppletive infinitive forms within one inflectional class (*maknuti/maći: maknem* 'to move', *zujati/zujiti: zujim* 'to buzz', *drhtati/drhtjeti: drhtim* 'to shiver', *grepsti/ grebati: grebem* 'to scratch' etc.) or two lexemes (having the same root) resulting in two inflectional classes (e.g. *šiti/šivati* 'to sew', *pakovati/pakirati* 'to pack', *bojati/bojiti* 'to colour', *crpsti/crpiti* 'to drain' etc.). Research into the aspects of Croatian verbal system might contribute to the understanding of overabundance in morphologically rich inflectional languages.

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

103

Sources

Birtić, Matea *et al.* (2012) *Školski rječnik hrvatskoga jezika*, Školska knjiga & Institut za hrvatski jezik i jezikoslovlje, Zagreb.

Parčić, Dragutin A. (1901 reprinted in 1995) *Rječnik hrvatsko-talijanski*, Artresor studio, Zagreb.

Rječnik hrvatskoga ili srpskoga jezika I–XXIII (1880–1976), Jugoslavenska akademija znanosti i umjetnosti, Zagreb.

Veliki Anić: *Rječnik hrvatskoga jezika* (2003) ed. Ljiljana Jojić, Novi Liber, Zagreb.

Rječnik hrvatskoga jezika (2000) eds. Jure Šonje, Anuška Nakić, Leksikografski zavod & Školska knjiga, Zagreb.

Croatian web corpus (hrWaC): http://nlp.ffzg.hr/resources/corpora/hrwac/

Croatian Language Corpus (Riznica): https://riznica.ihjj.hr

Croatian National Corpus: http://www.hnk.ffzg.hr/

References

Aronoff, Mark (1976) *Word Formation in Generative Grammar*, Linguistic Inquiry Monograph No. 1, MIT Press, Cambridge.

Babić, Stjepan *et al.* (1991) *Povijesni pregled, glasovi i oblici hrvatskoga književnoga jezika*, Nakladni zavod Globus & HAZU, Zagreb.

Barić, Eugenija *et al.* (1997) *Hrvatska gramatika*, Školska knjiga, Zagreb.

Bermel, Neil, Luděk Knittl (2012) "Corpus frequency and acceptability judgements: A study of morphosyntactic variants in Czech", *Corpus Linguistics and Linguistic Theory*, 8,2, 241–275.

Bermel, Neil, Luděk Knittl, Jean Russell (2015) "Morphological variation and sensitivity to frequency of forms among native speakers of Czech", *Russian Linguistics*, 39,105, 1–26.

Bošnjak Botica, Tomislava (2011) *Kategorija glagolske vrste u hrvatskom jeziku* ('*A category of verb type in Croatian*'), doctoral dissertation, Filozofski fakultet Sveučilišta u Zagrebu.

Clark, Eve V. (1987) "The principle of contrast: A constraint on language acquisition", *Mechanisms of language acquisition*, ed. Brian MacWhinney, Lawrence Erlbaum Associates, Hillsdale, NJ, 1–33.

Corbett, Greville G. (2007) "Canonical typology, suppletion and possible words", *Language*, 83, 8–41.

Corbett, Greville G. (2005) "The canonical approach in typology", *Linguistic Diversity and Language Theories*, eds. Zygmunt Frajzyngier, Adam Hodges and David S. Rood, Benjamins Amsterdam/Philadelphia, 25–49.

104
Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

Ćavar, Damir, Dunja Brozović Rončević (2012) "Riznica: the Croatian Language Corpus", *Prace filologiczne*, LXIII, 51–65.

Dahl, Östen (2004) "The growth and maintenance of linguistic complexity", *Studies in language companion series*, 71, Benjamins, Amsterdam.

Dressler, Wolfgang U. (1985) "Suppletion in word formation", *Historical semantics – historical word-formation*, ed. Jacek Fisiak, Mouton de Gruyter, Berlin, 97–112.

Dressler, Wolfgang. U. (1999) "Why collapse morphological concepts? *Behavioural and Brain Sciences*, 22, 1021.

Dressler, Wolfgang U. (2005) "Morphological Typology and First Language Acquisition: Some Mutual Challenges", *Morphology and Linguistic Typology*, *On-line Proceedings of the Fourth Mediterranean Morphology Meeting* (MMM4), eds. Geert Booij *et al.*, University of Bologna.

Fehringer, Carol (2004) "How Stable are Morphological Doublets? A Case Study of ə and Ø Variants in Dutch and German", *Journal of Germanic Linguistics*, 16, 4, 285–329.

Hippisley, Andrew, Marina Chumakina, Greville G. Corbett and Dunstan Brown (2004) "Constraints on suppletion", *Studies in Language*, 28, 387–418.

Hržica, Gordana (2012) "Daj mi to napisaj: preopćavanja glagolske osnove u usvajanju hrvatskoga jezika" ('Overregularisations of the verb stem in the acquisition of Croatian language'), *Suvremena lingvistika*, 38, 74, 189–208.

Jelaska, Zrinka (2005) *Hrvatski kao drugi i strani jezik*, Hrvatska sveučilišna naklada, Zagreb.

Kroch, Anthony S. (1989) "Reflexes of Grammar in Patterns of Language Change", *Language Variation and Change*, 1, 199–244.

Kroch, Anthony S. (1994) "Morphosyntactic variation", *Parasession on variation and linguistic theory*, *Papers from the 30th regional meeting of the Chicago Linguistics Society*, eds. Katherine Beals *et al.*, Chicago Linguistics Society, Chicago, 80–201.

Kuvač Kraljević, Jelena, Gordana Hržica (2015) *Croatian Adult Spoken Language Corpus*.

Lečić, Dario (2015) "Morphological doublets in Croatian" *Russian Linguistics*, 39, 375–393.

Ljubešić, Nikola, Filip Klubička (2014) "{bs,hr,sr}WaC — Web corpora of Bosnian, Croatian and Serbian", *Proceedings of the 9th Web as Corpus Workshop* (WaC-9), Gothenburg.

Ljubešić, Nikola, Tomaž Erjavec (2011) "hrWaC and slWac: Compiling Web Corpora for Croatian and Slovene", *Text, Speech and Dialogue 2011*, *Lecture Notes in Computer Science*, Springer.

Meľčuk, Igor (1994) "Suppletion: toward a logical analysis of the concept", *Studies in Language* 18, 2, 339–410.

Mörth, Karlheinz, Wolfgang U. Dressler (2014) "German plural doublets with and without meaning differentiation", *Morphology and meaning: Selected papers from*

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

105

the 15th International Morphology Meeting, Vienna, February 2012, eds. Franz Rainer *et al.*, John Benjamins, Amsterdam, 249–258.

Pinker, Steven (1984) *Language Learnability and Language Development*, Harvard University Press, Cambridge.

Sanders, Nathan C., Steven B. Chin (2009) "Phonological Distance Measures", *Journal of Quantitative Linguistics*, 16,1, 96–114. Author's manuscript available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2856103/.

Santilli, Enzo (2014) *Italian Comparatives: a Case of Overabundance?*, doctoral dissertation, University of Aquila.

Santilli, Enzo (2016) "The Puzzling Case of Overabundance in Italian Comparatives", *17th International morphology meeting: Competition in Morphology*. Vienna, 18–21 February 2016, poster presentation.

Silić, Josip, Ivo Pranjković (2005) *Gramatika hrvatskoga jezika*, Školska knjiga, Zagreb.

Slobin, Dan I. (1985) "Crosslinguistic evidence for the language-making capacity", *The Crosslinguistic Study of Language Acquisition 2, Theoretical Issues,* Lawrence Erlbaum Associates, Hillsdale, NJ, 1157–1256.

Stump, Gregory, Raphael Finkel (2013) *Morphological typology*, Cambridge University Press.

Tadić, Marko (2009) "New version of the Croatian National Corpus", *After Half a Century of Slavonic Natural Language Processing*, eds. Dana Hlaváčková *et al.*, Masaryk University, Brno, 199–205.

Thornton, Anna M. (2011) "Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology", *Morphological autonomy: perspectives from Romance inflectional morphology*, eds. Martin Maiden *et al*., Oxford University Press, 358–381.

Thornton, Anna M. (2012) "Reduction and maintenance of overabundance. A case study on Italian verb paradigms", *Word Structure* 5, 2, 457–468.

Thornton, Anna M. (2016) "Overabundance: A Canonical Typology", *17th International morphology meeting: Competition in Morphology*, Vienna, 18–21 February 2016, oral presentation.

Veselinova, Ljuba N. (2006) *Suppletion in Verb Paradigms: Bits and pieces of the puzzle*, Typological Studies in Language (TSL). John Benjamins.

Yarkoni, Tal, David Balota (2008) „Moving beyond Coltheart's N: A new measure of orthographic similarity", *Psychonomic Bulletin & Review*, 15, 5, 971–979. Available at: http://talyarkoni.com/papers/Yarkoni_Balota_Yap_PBR.pdf.

106

Tomislava Bošnjak Botica, Gordana Hržica, *Overabundance in Croatian dual-class verbs*
FLUMINENSIA, god. 28 (2016), br. 1, str. 83-106

SAŽETAK
Tomislava Bošnjak Botica, Gordana Hržica
PREOBILJE U MORFOLOGIJI: DVOVRSNI GLAGOLI U HRVATSKOME

Hrvatska se glagolska fleksija obično opisuje na temelju svrstavanja glagola u morfološke vrste, čiji je broj različit ovisno o pristupu, pri čemu je kriterij u klasifikaciji prisutnost ili odsutnost supletivnosti (promjene) i tip supletivnosti promjene u glagolskoj osnovi. Načelno, jedan glagol pripada samo jednoj vrsti (paradigmi), no postoje glagoli koji pripadaju dvjema vrstama, dakle, imaju dva paralelna niza osnova. Kod takvih se glagola nasuprot jednomu infinitivnom obliku nalaze dva prezentska u svim ćelijama (engl. *cells*) unutar vrste, to jest javlja se tzv. morfološko preobilje (engl. *overabundance*, Thornton 2012). Pritom je jedna od tih dviju paradigmi vrsta s kategorijskom supletivnošću (engl. *categorial suppletion*). Prezentska osnova vrste s kategorijskom supletivnošću ima veću fonološku udaljenost (engl. *phonological distance*) od infinitivne osnove nego prezentska osnova druge vrste. Drugim riječima, u oblikovanju prezentske osnove jedna je vrsta više, a druga manje transparentna. U uporabi se jedna vrsta uvijek češće rabi od druge, dakle, jedni oblici konkuriraju drugima. Ovim se istraživanjem nastoji prikazati preobilje oblika kod dvovrsnih glagola te odrediti može li se „natjecanje" između dvaju paralelnih oblika objasniti njihovom tendencijom da se priklone jednoj vrsti (koja bi se mogla smatrati prototipnom zbog čestotnosti i transparentnosti) ili drugim čimbenicima, točnije, fonološkom udaljenošću između dviju paradigma dvovrsnih glagola. Iz različitih je vrela – nekoliko rječnika hrvatskoga jezika i korpusa – izdvojeno više od 90 dvovrsnih glagola (u osnovnom obliku, a u prefigiranom kad nema osnovnoga). Otprilike ih je polovica potvrđena i u rječnicima i u nekom od korpusa. Ima i onih čija se dvovrsnost potvrđuje samo u korpusima, i to uglavnom u hrWaC-u. Ti glagoli u analizi čine zasebnu skupinu. Dok su neki dvovrsni glagoli češći u svojoj nesupletivnoj (transparentnijoj) vrsti, ostali pokazuju suprotno, što govori da na konkurenciju između dvaju oblika ne utječe samo pojednostavljenje paradigme. Kao predskazivači preferencije glagola prema jednoj ili drugoj vrsti promatrana su dva čimbenika: čestota i fonološka udaljenost. Ukupna je čestota glagola mjerena u najvećem dostupnom korpusu, hrWaC-u, pretragom lema. Fonološka je udaljenost za svaki glagol izračunata s pomoću Levenshteinove udaljenosti (Levenshtein 1965) na osnovi triju operacija: umetanja, brisanja i zamjene, kako je opisano u literaturi (Sanders and Chin 2009). Pokazalo se da je samo fonološka udaljenost čimbenik koji predviđa izbor vrste. Što je manja fonološka udaljenost između oblika dviju vrsta dvovrsnih glagola, to će se oni češće pojavljivati u oblicima transparentnije vrste. Rezultati upućuju na dva zaključka. Prvo, idu u prilog teoriji prirodne morfologije, po kojoj su neprirodne jezične pojave (uključujući supletivnost) češće u flektivnim jezicima. Nadalje, potvrđuju tvrdnju (Dressler 1999) da izrazito flektivni jezici često nemaju ili imaju tek slabi prototipni obrazac među konkurentskim morfološkim obrascima. Treba reći da u okviru ovoga teorijskoga okvira supletivnost spada među najmanje transparentne i prema tome najneprirodnije pojave (slaba je supletivnost samo jednu razinu niže na ljestvici u odnosu na jaku). To upućuje na pretpostavku da će se glagoli s većom fonološkom udaljenošću češće pojaviti u nesupletivnoj vrsti od glagola s manjom fonološkom udaljenošću, no dobiveni rezultati pokazuju suprotnu tendenciju. S obzirom na relativno malu količinu objašnjene varijance morat će se razmotriti dodatni čimbenici da bi se opisala varijabilnost u preferenciji dvovrsnih glagola. Potencijalni su kandidati čestota vrste, broj glagola po vrsti, čestota unutar ćelije, utjecaj fonotaktike i slično. Noviji se pristupi fonološkoj udaljenosti služe razlikovnim obilježjima fonema kao jedinicama u određivanju udaljenosti između dvaju dijelova (npr. Sanders i Chin 2009). Takva bi preciznija analiza mogla doprinijeti objašnjenju navedene pojave.

Ključne riječi: *preobilje oblika; morfološko natjecanje; glagolska fleksija; supletivnost; fonološka udaljenost; hrvatski jezik*