Simultaneous effects of inflectional paradigms and classes in processing of Serbian verbs

Dušica Filipović Đurđević & Isidora Gatarić

Department of Psychology, Faculty of Philosophy, University of Novi Sad, Serbia

In this paper we show that the processing of inflected verb forms is simultaneously influenced by the distributional properties of their inflectional paradigm (all the inflected forms of the given verb) and also by their inflectional class (all the verbs that conjugate in the same manner). Thus, we generalize a finding that was previously observed with nouns. We demonstrate that a divergence of the frequency distribution within inflectional paradigm from the frequency distribution within inflectional class (operationalized as Relative entropy between the two frequency distributions) is detrimental to processing. We present the results of a visual lexical decision experiment and the results of a simulation that was ran in the Naive Discriminative Reader, a simple computational model based on basic learning principles. We show that Relative entropy between an inflectional paradigm and an inflectional class predicts both empirically observed and simulated processing latencies. By doing so, we add to the body of research that investigates processing effects of information theory based descriptions of language. We also demonstrate that the effect of Relative entropy on the processing of morphology can arise as a consequence of the principles of discriminative learning in a system that maps input cues to outcomes, with no specification of morphology per se.

Key words: inflectional morphology, verbs, paradigms and classes, Relative entropy, Naive Discriminative Reader

Highlights:

- The typicality of inflected verbs frequency distribution expressed as Relative entropy.
- Relative entropy predicts the processing of inflected verb forms.
- The divergence of the inflectional paradigm from the class is detrimental to the processing.
- The observed pattern is simulated in the Naive Discriminative Reader.
- Morphological effects can arise as a consequence of simple learning principles.

Corresponding author: dfdurdevic@ff.uns.ac.rs

This paper is a continuation of a line of research that demonstrated that processing of nouns is affected both by the properties of frequency distribution of their inflected forms and the properties of summed (or average) frequency distribution of inflected forms of all the nouns that belong to the same inflectional class (Baayen, Milin, Filipović Đurđević, Hendrix, & Marelli, 2011; Milin, Filipović Đurđević, & Moscoso del Prado Martín, 2009). More precisely, it was shown that an increase in the divergence between the two distributions was followed by longer processing latencies and more errors in the visual lexical decision task. Here, we aimed to demonstrate that this finding could be generalized to other word classes, namely verbs. In other words, we wanted to demonstrate that verb processing was also more costly if the difference between the two distributions was large. Additionally, we aimed to demonstrate that this effect could arise as a consequence of some basic learning principles. We did so by running simulations in the Naive Discriminative Reader (NDR; Baayen et al., 2011), a model based on the principles of Naive Discriminative Learning (NDL) which had already been used to account for several morphological effects.

Information Theory and morphology research

Processing effects of variables based on information theory descriptions were demonstrated for many language-related phenomena (Ackerman, Blevins, & Malouf, 2009; Ackerman & Malouf, 2013; Baayen et al., 2011; Balling & Baayen, 2012; Filipović Đurđević 2007; Filipović Đurđević, Đurđević & Kostić, 2009; Frank, 2010, 2013; Hale, 2001, 2003, 2006; Kemps, Wurm, Ernestus, Schreuder, & Baayen, 2005; Kostić, 1991, 1995; Kostić & Havelka, 2002; Kostić, Marković, & Baucal, 2003; Kostić & Mirković, 2002; Levy, 2008; Milin, et al., 2009; Milin, Kuperman, Kostić, & Baayen, 2009; Moscoso del Prado Martin, Kostić, & Baayen, 2004; Pluymaekers, Ernestus, & Baayen, 2005; Tabak, Schreuder, & Baayen, 2005; Wurm, Ernestus, Schreuder, & Baayen, 2006). For example, in a series of research papers Kostić and his colleagues demonstrated that processing of inflected forms of nouns, adjectives, and verbs was influenced by the complexity of the given inflected form. In their research, this complexity was expressed as information load (surprisal; Shannon, 1948) derived from the average frequency per syntactic function/meaning of that form (i.e., its role in a sentence; Filipović & Kostić, 2003, 2004; Kostić, 1991, 1995; Kostić & Havelka, 2002; Kostić et al., 2003; Kostić & Mirković, 2002). The average frequency per number of syntactic functions/meanings of the inflected form in question was expressed relative to other inflected forms of the same lemma (stem), indicating the importance of the full inflectional paradigm. The influence of complexity of the full inflectional paradigm on processing was further demonstrated in a line of research that reported processing effects of the Inflectional Entropy (Baayen, Feldman, & Schreuder, 2006; Moscoso del Prado Martín et al., 2004; Tabak et al., 2005).

Furthermore, research that followed revealed that it was not only information complexity of the given inflected form, and information complexity

of the full set of related inflected forms that affected processing, but also the relation between the set of inflected forms of the given word (the inflectional paradigm) and the set of all the words which take the same exponents – build their inflected forms in the same way (inflectional class; Baayen et al., 2011; Hendrix, Bolger, & Baayen, 2016; Milin et al., 2009; for the same effects in derivational morphology see Kuperman, Bertram, & Baayen, 2010; Milin, Kuperman et al., 2009). The concept of inflectional paradigm and inflectional class was defined in linguistics (Anderson, 1992; Aronoff, 1994; Blevins, 2006). In psycholinguistic research the relation between the inflectional paradigm and the inflectional class was operationalized as Relative entropy, or the Kullback-Leibler divergence between the two frequency distributions. The frequency distribution of the paradigm consisted of surface frequencies of all inflected forms of the given noun, whereas frequency distribution of the class consisted of suffix (exponent) frequencies or cumulative frequencies of the given inflected forms of all the nouns that belong to the same class (take on the same exponents; e.g., feminine nouns). The larger the deviance of the paradigm frequency distribution from that of the class, the costlier the processing was. This indicated that words that had typical inflected form frequency distributions were easier to process than those with rare and atypical distributions of form frequencies.

The detrimental processing effect of dissimilarity between frequency distributions of the paradigm and the class was demonstrated in several different experimental settings. It had been first demonstrated in a visual lexical decision task by Milin et al. (2009; see also Milin, Kuperman et al., 2009), but it was soon also observed in a self-paced reading and primed lexical decision task (Baayen et al., 2011), as well as in an eye-tracking experiment (Kuperman et al., 2010). Baayen and his colleagues (2011) extended the finding to English prepositional phrases, showing that the (dis)similarity of the frequency distribution of the co-occurrence of the noun with different prepositions (i.e., the prepositional phrases of the given noun) and overall prepositional phrase frequency distribution (i.e., prepositional phrases of all the nouns) also influenced processing. Recently, the same effect was documented in a primed picture-naming task, where both effects on processing latencies and effects on ERPs were observed (Hendrix et al., 2016).

Naive discriminative learning perspective

The effects of relative entropy were simulated in the NDR (Baayen et al., 2011) – a computational model based on the principles of discriminative learning. This model entails a very simple architecture that maps orthographic input onto meaning (output). The input is specified as a set of unigrams, bigrams, or trigrams. For example, an inflected form of the word *gleda* (he looks) is segmented into bigrams *#g, gl, le, ed, da, a#*. The output refers to the meaning in general and is specified by the language unit that refers to the experience that is to be discriminated from other sets of experiences. This is usually achieved by using a lemma as the output unit (e.g., *gledati – to look*),

but the same can be achieved by using multi-word phrases. It is important to note that the output units are not interpreted as representational units per se, but rather as links between the orthographic input and the experience that the input is referring to. In order to separate the NDR output units from lexemes and other traditional language units, the authors termed the NDR output units as lexomes (Baayen, Shaoul, Willits, & Ramscar, 2015). In the course of training, the model learns to separate the cues that are good predictors of the output from the cues that are not good at predicting them. This is achieved by calculating the strengths of association between the cues and the outcomes. This process is based on the learning principle formulated by Rescorla and Wagner (1972). This principle involves a constant updating of cue-outcome associations in such a way that the strength of association is increased if both the cue and the outcome are present, whereas it is decreased if the cue is present without the outcome. In the absence of both the cue and the outcome the association strength remains unchanged. Since this process implies a constant updating, the system never reaches the equilibrium state. Therefore, in order to efficiently estimate cue-outcome association strengths, Danks (2003) proposed a set of equations which describe the equilibrium state of the system by introducing the assumption that in the stable state, all changes in association strengths equal zero. This reduced the estimation of association strengths to finding values which optimize conditional probabilities of the outcomes given the probabilities of the set of cues. Upon this estimation, the activation of the given outcome is calculated as the sum of association strengths of this outcome and all the cues that are present in the set of the input cues (e.g., all the bigrams that constitute a word). Finally, simulated reaction time is modelled as the inverse of this activation.

Even though morphology is not specified in the model, several morphological effects that were observed in various tasks were successfully simulated (Baayen et al., 2011; Hendrix, 2016; Hendrix et al., 2016). Although generally similar to the family of parallel distributed computational models regarding the unspecified, but emergent morphology (e.g., Gonnerman, Seidenberg, & Andersen, 2007; Plaut & Gonnerman, 2000; Rumelhart & McClelland, 1986; Seidenberg & Gonnerman, 2000; Seidenberg & Plaut, 2014), this model is of a more simplistic nature, as it does not include hidden layers and does not imply feedback activation. The core process in this model and the process that makes the learning possible and dynamic is the process of cue competition, as described by Ramscar, Yarlett, Dye, Denny, and Thorpe (2010).

The current study

In this study we firstly aimed to demonstrate that the findings of Milin et al. (2009) could be generalized to verbs. In other words, we aimed to show that processing of inflected forms of verbs was influenced by the divergence of the frequency distribution of the inflected forms of the given verb (its inflectional paradigm) from the summed frequency distribution of the inflected forms of all the verbs that are conjugated in the same way (its inflectional class). Secondly, we wanted to show that this effect could be simulated as a consequence of simple

learning principles, namely the principles of naive discriminative learning, as suggested by Baayen et al. (2011).

Serbian verbs. The morphology of Serbian verbs is characterized by several attributes. Depending on the syntactic role that they have in the sentence, verbs appear in different inflected forms. These forms denote several grammatical meanings. There are three persons (1st, 2nd, 3rd) and two numbers (sg, pl) denoting who performs the action. Also, there are three verb moods denoting the speaker's attitude (imperative, potential, and future II), and several verb tenses denoting the time of the action (e.g., present, future, etc.). Tenses are formed either by adding an inflectional suffix to the root (e.g., *Ja gleda-m – I look; Ti gleda-š – You look*), or by combining certain inflected forms of the verb with the appropriate inflected forms of auxiliary verbs (e.g., *Ja sam gleda-o – I looked; Ti si gleda-o – You looked*). Additionally, for some verb forms, agreement with the noun is denoted by grammatical number (e.g., *Devojčica je gleda-la – The girl looked; Devojčice su gleda-le – The girls looked*) and grammatical gender (e.g., *Devojčica je gleda-la – The girl looked; Dečak je gleda-o – The boy looked; Dete je gleda-lo – The child looked*). There are other verb-related grammatical attributes, such as aspect, which describes whether the verb denotes an action which is completed, the one which is still in progress, or the one which is repetitive. However, in the Serbian language, aspect is coded in the lexical meaning, not in inflectional morphology of the verb. The same holds for verb valence and/or transitivity, which describe the potential of the verb to attract other elements within a sentence.

Table 1 summarizes the inflected forms of Serbian verbs and their relations to grammatical categories of tense, person, number and gender (where applicable), whereas a simple list of inflected forms is presented in Table 2. For the purposes of illustration, we presented the inflected forms of verbs which have a root ending with phoneme /a/ (the remaining verb categories differ to some extent with respect to the distribution of suffixes, but the overall pattern is similar).

Table 1
Inflected forms of Serbian verbs that have a root ending with phoneme /a/, an example of verb gledati (to look): A) Inflected forms of simple tenses (those that include only verb's inflected form) for three persons, singular and plural; B) Inflected forms that construct complex tenses in combination with appropriate inflected forms of auxiliary verbs, for three genders, singular and plural; C) Inflected forms that do not agree with gender, number, or person, and are used to form certain tenses or syntactic phrases

A			Present	Imperative	Aorist	Imperfect	Future I
Singular	1st	ja (I)	gleda-m	/	gleda-h	gleda-h	gleda-ću
	2nd	ti (You)	gleda-š	gleda-j	gleda-ø	gleda-še	gleda-ćeš
	3rd	on/ona/ono (He/She/It)	gleda-ø	(neka) gleda	gleda-ø	gleda-še	gleda-ćeš
Plural	1st	mi (We)	gleda-mo	gleda-jmo	gleda-smo	gleda-smo	gleda-ćemo
	2nd	vi (You)	gleda-te	gleda-jte	gleda-ste	gleda-ste	gleda-ćete
	3rd	oni/one/ona (They)	gleda-ju	(neka) gleda-ju	gleda-še	gleda-hu	gleda-će

Table 1 (continued)

B		Active verbal adjective	Passive verbal adjective
Singular	Masculine	gleda-o	gleda-n
	Feminine	gleda-la	gleda-na
	Neuter	gleda-lo	gleda-no
Plural	Masculine	gleda-li	gleda-ni
	Feminine	gleda-le	gleda-ne
	Neuter	gleda-la	gleda-na

C	Infinitive	Present verbal adverb	Past verbal adverb
	gleda-ti	gleda-jući	gleda-vši

Relative entropy. For each inflected verb form two associated frequencies can be obtained: its surface frequency and the frequency of its inflectional suffix, as illustrated in Table 2. The frequency of an inflectional suffix is the cumulative frequency of the given inflected forms (e.g., inflected form ending with – *mo*, as in *gleda-mo*) of all the verbs that belong to the same class (e.g., *gleda-mo*, *peva-mo*, *zeva-mo*, etc.). The surface frequencies of all inflected forms of the given verb constitute the frequency distribution of the paradigm of that particular verb, whereas the suffix frequencies constitute the frequency distribution of the inflectional class of the given verb. As suggested by Milin et al. (2009), the divergence of the paradigm distribution from the class distribution could be operationalized as relative entropy or the Kullback-Leibler (KL) divergence ($D (p\|q)$; Cover & Thomas, 1991; Kullback, 1959):

$$D(p(x)\|q(x)) = \sum_{x} p(x)log\frac{p(x)}{q(x)}$$

$$= \sum_{i=1}^{n} f(w_i)/f(w)log\frac{f(w_i)/f(w)}{f(e_i)/f(e)}$$

(1)

Form frequency distribution of inflectional paradigm is denoted by *p(x)* in equation (1), whereas *q(x)* denotes the form frequency distribution of inflectional class. The inflected forms within each given distribution are denoted by *i*. The frequency of the *i*-th inflected form of the word *w* is marked by $f(w_i)$ and $f(w)$ stands for the cumulative frequency of all the inflected forms of the word *w* (stem frequency of *w*). The frequency of the *i*-th exponent, that is, the sum of the frequencies of *i*-th inflected forms of all the words from the given inflectional class are denoted by $f(e_i)$. Finally, *f(e)* stands for the cumulative frequency of all the inflected forms of all the words that belong to the given inflectional class. Table 2 illustrates how relative entropy for the verb *gledati* (to look) is calculated based on Equation (1).

Table 2

Inflected forms of verb gledati (to look) and their frequencies (per two millions), inflectional suffixes from its inflectional class, their frequencies (per two millions), and Relative entropy between the two distributions calculated based on Equation (1)

Inflected form	$f(w_i)$	$p(i)=f(w_i)/f(w)$	suffix	$f(e_i)$	$q(i)=f(e_i)/f(e)$	$p(i)log(p(i)/q(i))$
gleda-ti	58	32.90×10^{-3}	-ti	1746	53.70×10^{-3}	-6.99×10^{-3}
gleda-m	426	241.60×10^{-3}	-m	4976	152.90×10^{-3}	47.97×10^{-3}
gleda-š	92	52.20×10^{-3}	-š	1166	35.80×10^{-3}	8.51×10^{-3}
gleda-ø	374	212.10×10^{-3}	-ø	10891	334.70×10^{-3}	-42.02×10^{-3}
gleda-mo	43	24.40×10^{-3}	-mo	847	26.00×10^{-3}	$-.69\times10^{-3}$
gleda-te	8	4.50×10^{-3}	-te	264	8.10×10^{-3}	-1.15×10^{-3}
gleda-ju	197	111.70×10^{-3}	-ju	3679	113.10×10^{-3}	$-.58\times10^{-3}$
gleda-j	116	65.80×10^{-3}	-j	1149	35.30×10^{-3}	17.77×10^{-3}
gleda-jmo	1	$.60\times10^{-3}$	-jmo	41	1.30×10^{-3}	$-.20\times10^{-3}$
gleda-jte	21	11.90×10^{-3}	-jte	282	8.70×10^{-3}	1.64×10^{-3}
gleda-h	17	9.60×10^{-3}	-h	227	7.00×10^{-3}	1.35×10^{-3}
gleda-še	5	2.80×10^{-3}	-še	69	2.10×10^{-3}	$.36\times10^{-3}$
gleda-smo	1	$.60\times10^{-3}$	-smo	27	$.80\times10^{-3}$	$-.09\times10^{-3}$
gleda-ste	.1	$.10\times10^{-3}$	-ste	7	$.20\times10^{-3}$	$-.03\times10^{-3}$
gleda-hu	6	3.40×10^{-3}	-hu	28	$.90\times10^{-3}$	2.03×10^{-3}
gleda-o	175	99.20×10^{-3}	-o	2702	83.00×10^{-3}	7.68×10^{-3}
gleda-la	52	29.50×10^{-3}	-la	1104	33.90×10^{-3}	-1.80×10^{-3}
gleda-lo	9	5.10×10^{-3}	-lo	376	11.60×10^{-3}	-1.81×10^{-3}
gleda-li	50	28.40×10^{-3}	-li	1257	38.60×10^{-3}	-3.81×10^{-3}
gleda-le	17	9.60×10^{-3}	-le	266	8.20×10^{-3}	$.69\times10^{-3}$
gleda-n	2	1.10×10^{-3}	-n	128	3.90×10^{-3}	$-.61\times10^{-3}$
gleda-na	1	$.60\times10^{-3}$	-na	122	3.70×10^{-3}	$-.47\times10^{-3}$
gleda-no	6	3.40×10^{-3}	-no	110	3.40×10^{-3}	$.01\times10^{-3}$
gleda-ni	.1	$.10\times10^{-3}$	-ni	45	1.40×10^{-3}	$-.08\times10^{-3}$
gleda-ne	2	1.10×10^{-3}	-ne	40	1.20×10^{-3}	$-.04\times10^{-3}$
gleda-jući	62	35.20×10^{-3}	-jući	473	14.50×10^{-3}	13.49×10^{-3}
gleda-vši	.1	$.10\times10^{-3}$	-vši	16	$.50\times10^{-3}$	$-.05\times10^{-3}$
gleda-ću	10	5.70×10^{-3}	-ću	109	3.30×10^{-3}	1.30×10^{-3}
gleda-ćeš	4	2.30×10^{-3}	-ćeš	48	1.50×10^{-3}	$.42\times10^{-3}$
gleda-će	8	4.50×10^{-3}	-će	294	9.00×10^{-3}	-1.36×10^{-3}
gleda-ćemo	.1	$.10\times10^{-3}$	-ćemo	38	1.20×10^{-3}	$-.07\times10^{-3}$
gleda-ćete	.1	$.10\times10^{-3}$	-ćete	13	$.40\times10^{-3}$	$-.05\times10^{-3}$
$f(w)=1763.5$				$f(e)=32540$		$D(p\|\|q)=.041$

Experiment

We presented the inflected verb forms in a visual lexical decision task in order to demonstrate that their processing was affected by relative entropy between the distribution of frequencies of the inflected forms of the particular verb and the distribution of cumulative frequencies of the inflected forms of all

the verbs that conjugate in the same manner (calculated based on Equation 1). We predicted that, as observed with nouns (Milin et al., 2009), Relative entropy would be positively correlated with processing latencies, indicating that verbs with more prototypical distributions of inflected form frequencies are less costly to process than the verbs with more deviant frequency distributions.

Method

Participants. Sixty-nine students from the Department of Psychology at the Faculty of Philosophy in University of Novi Sad took part in the experiment. They were all native speakers of Serbian with normal or corrected-to-normal vision, who reported of no reading difficulties. The participants were randomly assigned to one of four groups based on the list of the stimuli with which they were to be presented.

The research was approved by the Ethical Committee of the Department of Psychology at the Faculty of Philosophy in University of Novi Sad. All participants signed the informed consent form.

Materials and design. We selected 105 verbs of type V and 152 verbs of type VI, as categorized in the grammars of Serbian language (Stanojčić & Popović, 1992). Type V verbs were those with a stem ending with /a/ (e.g., *pevati* – to sing) and type VI verbs were those with a stem ending with /i/ (e.g., *raditi* – to work). The verbs were selected based on the Frequency Dictionary of Contemporary Serbian Language (Kostić, 1999). The selection criterion for a verb to be included in the stimuli list was for it to have as many as possible inflected forms with non-zero frequency.

The verbs were presented in two different inflected forms of the present tense: the 1st person singular (*pevam* – I sing; *radim* – I work) and the 2nd person singular (*pevaš* – You sing; *radiš* – You work). All of the verbs appeared in both of these forms, but not within the same list: the four lists were formed in such a way that all of the verbs within a list had the same inflectional suffix (as illustrated in Table 3; see Appendix A for the full list of stimuli).

We created pseudo words that mirrored the structure of type V and type VI words and within each group we created lists of pseudo words that ended with the same inflectional suffixes as the verbs from that particular list.

Table 3
Structure of stimuli that were presented in the experiment – each list was presented to different group of participants

Type V verbs		Type VI verbs	
1st person singular	2nd person singular	1st person singular	2nd person singular
List 1	List 2	List 3	List 4
PEVAM	PEVAŠ	RADIM	RADIŠ
(I sing)	(You sing)	(I work)	(You work)

The selection of stimuli and of the design complied with our attempt to make our study comparable to that of Milin and colleagues (2009). Along the same line, each participant was presented with a single list of stimuli, that is, with the verbs and pseudo words that ended with the same inflectional suffix.

The dependent variable was response latency and the crucial predictor was the KL divergence between probability distribution of the inflected forms of the given verb (inflectional paradigm) and the summed distribution of the verb's class (either type V or type VI verbs). The additional independent variables were the order of trial presentation, word length in letters, (log) surface frequency, and (log) lemma frequency.

Procedure. Open Sesame experimental software (Mathôt, Schreij, & Theeuwes, 2012) was used for stimuli presentation and data recording. We presented the participants with a visual lexical decision task. Each trial started with a blank screen that was presented for 500ms, followed by a fixation point presented for 1000ms. After that, the target stimulus would appear and would remain on the screen until the participant responded, or until the 1500ms timeout. The responses were to be given by mouse-button press, with *yes* being mapped onto the index finger pressing the left mouse button (the dominant action), and *no* being mapped onto the middle finger pressing the right mouse button. The order of stimuli presentation was randomized separately for each participant. Prior to the main session, the participants were presented with 10 practice trials in order to become familiarized with the task. The stimuli presented in the practice session were not presented in the main session and were not included in the analyses.

Results and discussion

Prior to the analysis, we excluded one participant and 11 verbs with above 25% error-rate. The overall error-rate was low – 9.4% on average.

The reaction times were log transformed in accordance with the recommendations described in Baayen and Milin (2010). In order to control for collinearity, all of the predictors were centred to zero and divided by standard deviation (Gelman & Hill, 2007). Collinearity in our dataset was close to high (28.38), as tested using the Kappa coefficient (Belsley, Kuh, & Welsch, 1980; see Appendix B for pair-wise correlations of predictors). Therefore, we analysed the data using the *mgcv* package (Wood, 2006, 2011) in R statistical software (R Core Team, 2015). We fitted mixed-effect generalized additive models to processing latencies as this analysis is less sensitive to collinearity in the set of predictors. In order to avoid the possibility of the coefficients being influenced by extreme values, the models were refitted after excluding points with residuals that exceeded the range of ±2.5 standard units. As this procedure did not bring any substantial change in the structure of the results, we reported the coefficients from the refitted model. In order to test whether inclusion of the given parameter in the model was justified by the data, we employed the *itsadug* package (van Rij, Wieling, Baayen, and van Rijn, 2015) to compare different variants of the model. This was achieved by testing the differences between the values of maximum likelihood (ML) as the measure of goodness of fit (AIC scores were not compared, as our model included component that accounted for autocorrelation in the data, in which case AIC scores are not reliable, as noted by the authors of the package). In the process of ML scores comparisons, we took into account statistical significance of Chi-square. In addition to that, as suggested by the authors of the package, we also considered the ML difference values (the difference in scores which is less than five is considered too small). We tested for the effects of the order of trial presentation, verb type, suffix, word length in letters, (log) surface frequency, (log) lemma frequency, relative entropy, as well as random effects of participants and items. The coefficients from the final model are presented in Table 4.

Table 4

Coefficients from the generalized additive model fitted to response latencies

Parametric coefficients:	Estimate	*Std. Error*	*t*	*Pr(>|t|)*
Intercept	6.52	.02	348.53	.000
Trial (order of presentation)	.01	.01	.89	.37
Word length (in letters)	.02	.004	5.27	.000
Smooth terms:	edf	ref.df	*F*	*p*
Tensor product smooth for Relative entropy and (log) Lemma frequency	3.001	3.001	22.47	.000
by-Participant factor smooths for Trial (order of presentation)	233.17	611.00	8.91	.000
by-Item random intercept	162.59	243.00	2.02	.000

The best model that was justified by the data was the one that included order of trial presentation, word length in letters, interaction between (log) lemma frequency and relative entropy, as well as by-participant smooths for the order of trial presentation, and by-item random intercept. As expected, an increase in word length was followed by longer processing latencies. Along the same line, both (log) lemma frequency and relative entropy affected processing in the expected direction – it took less time to process words of higher frequency and it took more time to process words with larger relative entropies. Additionally, there was an interaction between (log) lemma frequency and relative entropy that revealed joint effect of the two predictors. As can be seen in Figure 1, the fastest responses were observed for words of high frequency and low relative entropy (dark shade in the graph), whereas the response time showed a tendency to become longer as the values of frequency decreased and the values of relative entropy increased (lighter shades in the graph). Model comparisons revealed that this interaction was justified by the data, because when taking the number of the parameters in the model into account, the fit achieved by this model (as indicated by ML values) was supreme both compared to the model with separate linear effects of (log) lemma frequency and relative entropy ($ML_{difference}$ = 10.427, $AIC_{difference}$ =14.85, χ^2 = 10.427, p = .0003), and the model with separate smooths for the two predictors in question ($ML_{difference}$ = 6.441, $AIC_{difference}$ = 7.95, χ^2 = 6.441, p = .0003).
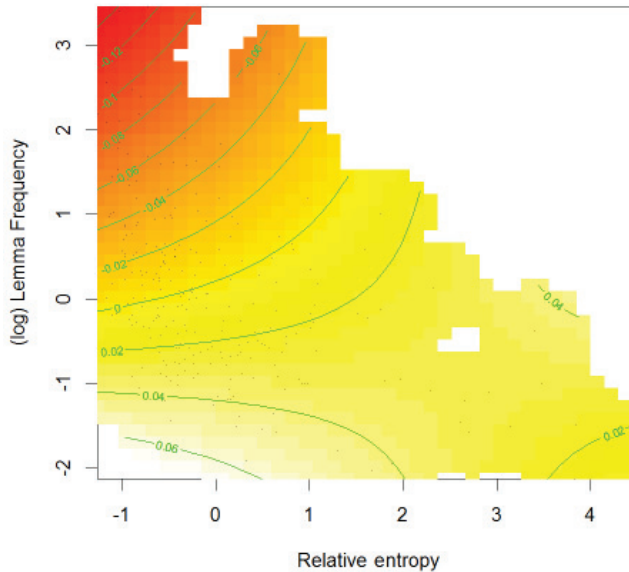
Figure 1. Generalized additive model fitted to log transformed response latencies: partial effects of relative entropy by (log) lemma frequency interaction; all variables are scaled; lighter shade denotes longer reaction time.

Our results showed that verbs with atypical distribution of inflected form frequencies were processed more slowly than verbs with inflected form frequency distribution that resembled the average distribution of the inflectional class to which they belong. Thus, we demonstrated that findings of Milin and colleagues (2009) could be generalized to verbs.

Simulation

In the next step, we tried to demonstrate that such effect could appear as a consequence of patterns that arise from mapping of orthography to semantics, without explicit coding of morphological units. We did so by running simulation in the Naive Discriminative Reader (Baayen et al., 2011), a simple computational model based on the Rescorla-Wagner equations (Rescorla & Wagner, 1972), which we described earlier.

Method

The simulation was performed using the R statistical software (R Core Team, 2015) and the *ndl* package (Arppe et al., 2015), the details of which are described in Baayen et al. (2011).

In order to build the database used for the model, for every lemma that was included in the study (257 in total: 105 of Type V and 152 of Type VI), we selected its inflected forms which were found in the Frequency Dictionary of Contemporary Serbian Language (Kostić, 1999) and their surface frequencies. In total, our database consisted of 4385 words and their corresponding frequencies. In the first step, every word was split into bigrams, and these were used as cues in the model, whereas the lemmas were treated as outcomes. Association strengths were estimated for every cue-outcome pair. Based on them, we calculated activations for every lemma that was provided by each individual inflected form (i.e., the set of cues that were contained in that form). Finally, we modelled simulated reaction time as the negative value of the association strength. Earlier simulations in the NDR revealed that including some additional information improved the overall fit of simulated data to empirically observed data (Baayen et al., 2011). However, our goal here was not to demonstrate the full power of the model in question, but to demonstrate that the model can account for the effects of relative entropy that was observed empirically. Therefore, we opted for the simplest approach in modelling simulated RT. Finally, we log-transformed simulated RT in order to correct the skewness of the distribution (in order to be able to apply the log-transformation, we added 1.1 to each simulated RT, thus making all the values fit in the range of the values to which the log function can be applied – i.e., above zero).

Results and discussion

We first calculated the Pearson correlation coefficient for the empirically observed response latencies to every inflected form of the verb and their simulated variants. Our results revealed a moderate, but significant positive correlation between the two: $r = .13$, $t(512) = 3.035$, $p = .003$. The observed coefficient was smaller than the one reported earlier in similar simulations (Baayen et al., 2011). However, this was not surprising as we tentatively applied the simplest possible way to model simulated RT, whereas the earlier simulations included many additional parameters to improve the fit.

Next, we fitted the linear regression model to (log) simulated reaction times in order to test whether the effects of the predictors that accounted for empirical reaction times could be observed here as well. The coefficients from the model are presented in Table 5.

Table 5
Coefficients from the linear model fitted to simulated reaction times

	estimate	Std. Error	t	Pr(>|t|)
Intercept	-.93	.03	-27.77	.000
Word length in letters	.07	.03	2.22	.027
(log) Lemma frequency	-.28	.04	-7.77	.000
Relative entropy	.11	.04	2.79	.005
(log) Lemma frequency * Relative entropy	.08	.03	2.61	.009

The model fitted to simulated RT revealed very similar effects as the one fitted to the empirically observed processing latencies. Firstly, as illustrated in Figure 2, we observed that word length in letters had an inhibitory effect on simulated RT, as well as in the case of empirical data. Along the same line, (log) lemma frequency had a facilitatory influence on both observed and simulated RT.
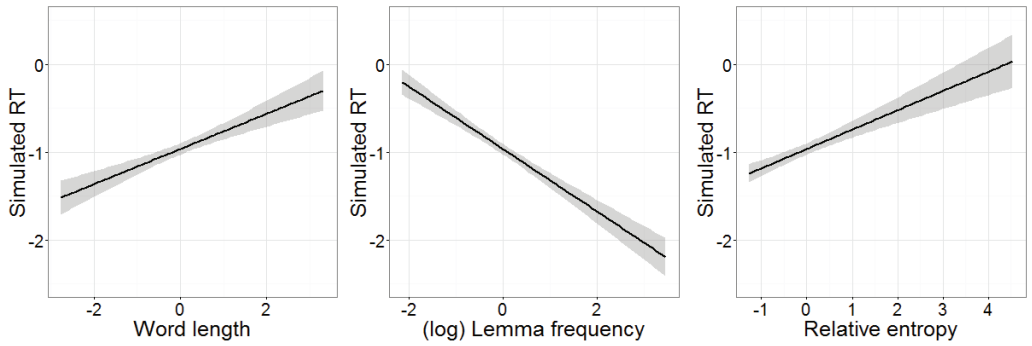
Figure 2. The effects of word length, (log) lemma frequency, and relative entropy on reaction times simulated by NDR.

Crucially, the effect of relative entropy was significant here, as well. As with empirical data, an increase in relative entropy was followed by an increase in simulated RT. The interaction between (log) lemma frequency and relative entropy was also significant in the case of simulated data, as was the case with empirical data. As illustrated in Figure 3 (produced by using loess and predict function for the purposes of illustration), the joint effect of (log) Lemma frequency and Relative entropy was similar to that observed on empirical data and illustrated in Figure 1.
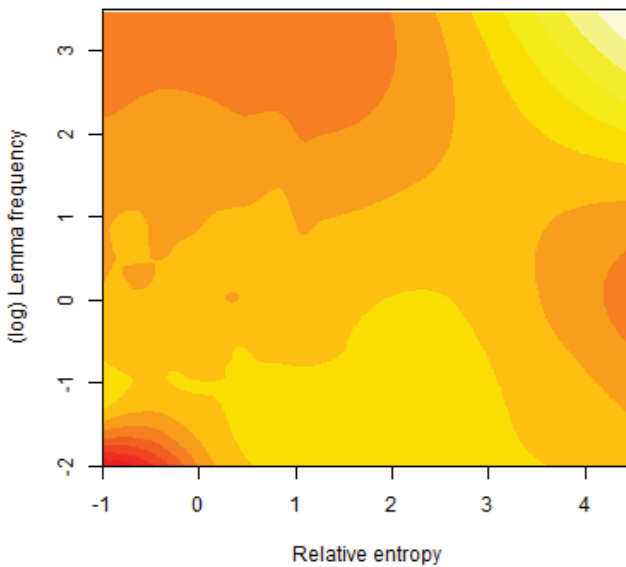

Figure 3. Interaction of (log) lemma frequency and relative entropy effects on (log) reaction times simulated by NDR (all predictors are scaled); lighter shade denotes longer simulated reaction time.

We therefore demonstrated that the detrimental effect of the divergence from the prototypical distribution of inflected form frequencies could emerge as a consequence of orthography-to-semantics mapping in a model that has no explicit representation of morphology. We tried to keep the model as simple as possible, which inevitably led to many limitations, and consequently low value of the correlation coefficient. Firstly, we introduced lemma (or stem) as the shortcut to semantics of the word. This was done for matters of simplicity, not because we argued that semantic representations are local by nature. On the contrary, we believe that they are highly distributed, as suggested by numerous research studies (for a review see McRae & Jones, 2013; Meteyard, Rodriguez Cuadrado, Bahrami, & Vigliocco, 2012), and that introducing this into the model would improve the fit to the empirical data. Additionally, when calculating simulated RT, we took into account only the inverse activation strength for the inflected form in question, thus ignoring its orthographic neighbours and other factors that could improve the fit.

General discussion

Our results revealed that processing of inflected verbs was affected by relation of two frequency distributions to which they simultaneously belong: the distribution of surface frequencies of all inflected forms of the given verb (i.e., its inflectional paradigm) and the distribution of frequencies of the exponents which build those inflected forms (i.e., its inflectional class). The divergence of paradigm frequency distribution from that of the class, as operationalized through Relative Entropy (or the Kullback-Leibler divergence) affected processing in such a way that verbs with divergent paradigm distributions (i.e., high Relative entropy) took more time to process. In other words, verbs with atypical frequency distribution of their inflected forms were more demanding in terms of processing.

This result is a direct continuation of the findings of Milin et al. (2009), who observed the same for nouns. It is also in accordance with the results of several other investigations that demonstrated similar effects in different tasks and different stimulus presentation conditions. For example, Nenadić, Tucker, and Milin (2016) observed the same effect in an auditory lexical decision task. Importantly, these authors presented participants with a list of mixed inflected forms, thus demonstrating that the effect of relative entropy as observed in Milin et al. (2009) and in our experiment was not an artefact of processing strategy developed by participants being presented with a single inflected form. Baayen et al. (2011) reported of the similar effect in a primed visual lexical decision and self-paced reading task. Here, processing of the inflected target noun was affected by the Weighted Relative entropy between the form-frequency distribution of the target noun and its preceding prime. The same authors calculated the Relative entropy for English nouns by looking at the distribution of prepositional phrases. The frequency distribution of the noun occurring with different prepositions was considered as its paradigm distribution, and general frequency distribution

of prepositions occurring with all the nouns was considered as its class. The Relative entropy between the two distributions was again positively correlated with processing time. Recently, Hendrix et al. (2016) documented the effect of Relative entropy on ERP measures as well. They observed that theta range oscillations (with the temporal onset comparable to that of word frequency and mostly prominent in parietal and occipital areas) were of greater amplitude for higher levels of Relative entropy, thus indicating that words with atypical paradigmatic frequency distributions demand additional processing. Some recent investigations demonstrated that the effect of Relative entropy could be generalized to adjectives as well, although some additional considerations need to be taken into account (Filipović Đurđević & Milin, 2018). For example, in the case of adjectives, the relevant unit of frequency distribution was not the frequency of inflected form, but of the grammatical case. This could be related to the fact that adjectives are usually followed by nouns and that adjectival meanings and syntactic role are fully realized in an adjective-noun phrase. In the field of derivational morphology effects of Relative entropy were documented by Milin, Kuperman et al. (2009).

None of the traditional models of processing of morphologically complex words could account for the observed effect of relative entropy. Both decomposition models (Taft & Forster, 1975; Taft 1979, 1994), and full-form models (Manelis & Tarp, 1977; Butterworth, 1983) address the characteristics of the individual inflected form and therefore make no predictions neither regarding the effects of the full set of inflected forms of the word, nor the effects of the relation between the distributions of paradigm and class.

As stated by Hendrix et al. (2016), the effect of Relative entropy between the two distributions could be fit in the exemplar-based approach. However, this would place a heavy load on the cognitive system either in terms of storage (very high number of exemplars due to a large number of rare events in language, as explained by Baayen, Hendrix, & Ramscar, 2013) or in terms of processing (on-line calculating distances between distributions). A much simpler explanation of the Relative entropy effect arises through the principles of discriminative learning (Arnon & Ramscar, 2012; Ramscar et al., 2010; Ramscar, Dye, & Clein, 2013; Rescorla & Wagner, 1972). These principles are integrated in the Naive Discriminative Reader model (NDR; Baayen et al., 2011) – a simple computational model that only searches for the optimal way of mapping the structure of the input layer to that of the output layer by applying the equations of Rescorla and Wagner (1972), and the equilibrium equations of Danks, 2003). We therefore ran the simulations by using the NDR computational model and observed a significant correlation coefficient between the observed and the simulated reaction times. Additionally, the regression analyses revealed fully comparable pattern of results for observed and simulated reaction time. As with empirical data, Relative entropy was positively correlated with simulated reaction time. This added to the body of evidence that the effect of Relative Entropy can arise as a consequence of the basic learning principles (Baayen et al., 2011; Hendrix et al., 2016).

It is important to note that, in naive discriminative learning approach the accounting for the Relative entropy effect comes for free. We did not feed the model with any of the information that would point to relation between the two distributions (paradigm and class). The model was trained only on mappings of word's bigrams to its lemma. In this model, the effect of Relative entropy appears simply as a consequence of the way the distributional properties of the language system "shape the associations between orthographic input cues and semantic outcomes across sequences of words" (Hendrix et al, 2016, p. 3). As inflected form of the word marks the syntactic role of the given word in the sentence, it entails in itself the syntactic context in which it occurs. Therefore, the mapping of different inflected forms to their common semantic outcome implies the mapping of different contexts in which they occur. Thus, this process is shaped by the distributional properties of the language system and its relation to the experience with the world. According to discriminative learning perspective, the goal of language use is to discriminate among the many cues in the linguistic input those that are good at predicting semantic outcomes from those that are not predictive. Here, semantic outcomes denote experiences with the world that linguistic cues are referring to. It could also be stated that discriminative learning operates by minimizing the uncertainty between the set of cues and the set of outcomes (Baayen et al., 2013). This point of view is very similar to the one taken by Information Theory (Shannon, 1948). As Ramscar and Port (2016) state, both Information Theory and Discriminative Learning deal with reduction of uncertainty about the current state of the system among all the possible states that the system could be in.

Although we primarily interpreted our findings in the light of discriminative learning approach (Baayen et al., 2011), it is important to note that our goal was not to strictly point to this approach as the single possible explanation for the observed pattern. Our goal was to point out that morphological effects can be accounted for without explicit introduction of morphological units. In that matter, our results would fit well with the broad category of connectionist models, which have already been applied in accounting for various similar effects both inside (Joanisse & Seidenberg, 1999; Plaut & Gonnerman, 2000; Woollams, Joanisse, & Patterson, 2009) and outside of morphology (Cree & McRae, 2003; Harm & Seidenberg, 1999, 2004; Seidenberg & McClelland, 1989). For example, Mirković, Seidenberg, and Joanisse (2011) demonstrated that the connectionist network can simulate a complex pattern of Serbian noun inflection. Unlike the performance of the rule-based system, the performance of the connectionist network that was trained to generate inflected forms was sensitive to the same variables that influenced the behaviour of native speakers. One such variable was inflectional neighbourhood size – "the proportion of items in a corpus behaving in the same way across inflectional forms (friends)" (Mirković et al., 2011, p. 663). This variable is highly similar to the concept of inflectional class, as we applied it. The basic difference between the two is that inflectional neighbourhood is empirically established, whereas the inflectional class is adopted from grammars. Although there is much overlap between the words that would belong to each of the two sets, there are some potential

differences, as some inflectional classes could be divided into several inflectional neighbourhood sets (e.g., due to slight changes to the stems). However, Mirković and colleagues (2011) examined the effects of the *size* of such set, whereas we looked into the *relation* between the distribution of the average probabilities of inflected forms from the set of inflectional "friends" and the distribution of the inflected form probabilities of the particular word. Additionally, the two studies did not apply the same task (inflectional form generation vs. visual lexical decision). Therefore, although the two studies converge in general conclusions, further research would be needed to fully compare the concepts they applied.

Conclusion

In this paper we brought two main insights: we generalized a previously reported empirical effect and demonstrated that this effect could arise as a consequence of simple learning principles. We reported the results of one experiment and one simulation.

The results of the experiment demonstrated that processing of inflected verbs is affected by the typicality of the frequency distribution of all inflected forms of that particular verb (i.e., by the divergence between frequency distribution of inflectional paradigm from frequency distribution of the inflectional class). This effect has previously been demonstrated for nouns (Milin et al., 2009), and our paper generalized the finding to a novel word class, the one that does not follow nominal inflections, but another type of inflection (conjugation). It therefore revealed that universal principle is applicable to different manifestations of inflectional morphology.

The results of the simulation revealed that observed reaction times could be simulated by a simple computational model based on the principles of discrimination learning (NDR; Baayen et al., 2011). It therefore added to the body of evidence that demonstrate how very complex morphological phenomena can naturally occur in a system with no explicit representation of morphology as a consequence of simple and universal learning principles.

References

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In James P. Blevins, & Juliette Blevins (Eds.), *Analogy in Grammar: Form and Acquisition*. Oxford University Press, Oxford, 54–82.

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language, 89*, 429–464. https://doi.org/10.1353/lan.2013.0054

Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511586262

Aronoff, M. (1994). *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, MA: MIT Press.

Arnon, I., & Ramscar, M. (2012) Granularity and the acquisition of grammatical gender: How order of acquisition affects what gets learned. *Cognition, 122*(3), 292–305. https://doi.org/10.1016/j.cognition.2011.10.009

Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., & Shaoul, C. (2015). *ndl: Naive Discriminative Learning. R package version 0.2.17*. http://CRAN.R-project.org/package=ndl

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55*(2), 290–313. https://doi.org/10.1016/j.jml.2006.03.008

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech, 56*, 329–347. https://doi.org/10.1177/0023830913484896

Baayen, R. H., & Milin, P (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28. https://doi.org/10.21500/20112084.807

Baayen, R. H., Milin, P., Filipović Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*, 438–482. https://doi.org/10.1037/a0023851

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2015). Comprehension without segmentation: A proof of concept with naive discrimination learning. *Language, Cognition, and Neuroscience*, *31*(1), 106–128.

Balling, L., & Baayen, R. H. (2012) Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition, 125*, 80–106. https://doi.org/10.1016/j.cognition.2012.06.003

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*, Wiley Series in Probability and Mathematical Statistics, New York. https://doi.org/10.1002/0471725153

Blevins, J. P. (2006). English inflection and derivation. In B. Aarts, & A. M. McMahon (Eds.), *Handbook of English linguistics* (pp. 507–536). London: Blackwell.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language Production, Vol. II: Development, writing and other language processes*. London: Academic Press.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons. https://doi.org/10.1002/0471200611

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132*, 163–201. https://doi.org/10.1037/0096–3445.132.2.163

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology, 47*(2), 109–121. https://doi.org/10.1016/S0022–2496(02)00016–0

Filipović, D., & Kostić, A. (2003). Kognitivna obrada prideva [Cognitive processing of adjectives]. *Psihologija*, *36*(3), 353–378. doi:10.2298/PSI0303353F

Filipović, D., & Kostić, A. (2004). Kognitivni status roda prideva u srpskom jeziku [Cognitive status of adjectives' gender in Serbian language]. *Psihološka istraživanja*, *14*, 125–168.

Filipović Đurđević, D. (2007). *Polysemy effect in processing of Serbian nouns* (Unpublished doctoral thesis). University of Belgrade, Belgrade, Serbia.

Filipović Đurđević, D., Đurđević, Đ., & Kostić, A. (2009). Vector based semantic analysis reveals absence of competition among related senses. *Psihologija, 42*(1), 95–106. https://doi.org/10.2298/PSI0901095F

Filipović Đurđević, D., & Milin, P., (2018). Information and Learning in Processing of Adjective Inflection, *Cortex*. https://doi.org/10.1016/j.cortex.2018.07.020

Frank, S. L. (2010). Uncertainty reduction as a measure of cognitive processing effort. *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics* (pp. 81–89). Uppsala, Sweden: Association for Computational Linguistics.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science, 5*, 475–494. https://doi.org/10.1111/tops.12025

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General, 136*, 323–45. https://doi.org/10.1037/0096–3445.136.2.323

Hale, J. (2001). A probabilistic Early parser as a psycholinguistic model. In *Proceedings of the second conference of the North American chapter of the Association for Computational Linguistics, volume 2*, (pp. 159–166). Pittsburgh, PA: Association for Computational Linguistics.

Hale, J. (2003). The information conveyed by words. *Journal of Psycholinguistic Research*, *32*, 101–123. https://doi.org/10.1023/A:1022492123056

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*, 643–672. https://doi.org/10.1207/s15516709cog0000_64

Harm, M. W., & Seidenberg, M.S. (1999) Phonology, Reading Acquisition, and Dyslexia: Insights from Connectionist Models. *Psychological Review, 106*(3), 491–528. https://doi.org/10.1037/0033–295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review, 111*(3), 662–720. https://doi.org/10.1037/0033–295X.111.3.662

Hendrix, P. (2016). *Experimental explorations of a discrimination learning approach to language processing* (Unpublished doctoral dissertation). Eberhard Karls Universität, Tübingen, Germany.

Hendrix, P., Bolger, P., & Baayen, R. (2016). Distinct erp signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America*, 7592–7597. https://doi.org/10.1073/pnas.96.13.7592

Kemps, R., Wurm, L., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005) Prosodic cues for morphological complexity: Comparatives and agent nouns in Dutch and English. *Language and Cognitive Processes, 20*, 43–73. https://doi.org/10.1080/01690960444000223

Kostić, A. (1991). Informational approach to processing inflected morphology: Standard data reconsidered. *Psychological Research, 53*(1), 62–70. https://doi.org/10.1007/BF00867333

Kostić, A. (1995). Informational load constraints on processing inflected morphology. U L. B. Feldman (Ed.), *Morphological Aspects of Language Processing*. New Jersey: Lawrence Erlbaum, Inc., Publishers.

Kostić, A., & Havelka, J. (2002). Processing of a verb tense. *Psihologija, 35*(3–4), 299–316. https://doi.org/10.2298/PSI0203299K

Kostić, A., Marković, T., & Baucal, A. (2003). Inflected morphology and word meaning: orthogonal or co-implicative cognitive domains? In H. Baayen, & R. Schreuder (Eds.). *Morphological Structure in Language Processing* (pp. 1–44). Berlin: Mouton de Gruyter.

Kostić, A., & Mirković, J. (2002). Processing of inflected nouns and levels of cognitive sensitivity. *Psihologija, 35*(3–4), 287–299. https://doi.org/10.2298/PSI0203287K

Kostić, Đ. (1999). *Frekvencijski rečnik savremenog srpskog jezika* [Frequency dictionary of modern Serbian language]. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju.

Kullback, S. (1959). *Statistics and Information Theory*. New York: J. Wiley and Sons.

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language, 62*, 83–97. https://doi.org/10.1016/j.jml.2009.10.001

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Manelis, L. & Tarp, D., (1977). The processing of affixed words. *Memory & Cognition, 5*, 690–695. https://doi.org/10.3758/BF03197417

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428–011-0168–7

McRae, K., & Jones, M. N. (2013). Semantic memory. In D. Reisberg (Ed.), The Oxford Handbook of Cognitive Psychology (pp. 206–219). Oxford, UK. https://doi.org/10.1093/oxfordhb/9780195376746.013.0014

Meteyard, L., Rodriguez Cuadrado, S., Bahrami, B., & Vigliocco, G. (2012) Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Milin, P., Filipović Đurđević, D., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language, 60*, 50–64. https://doi.org/10.1016/j.jml.2008.08.007

Milin, P., Kuperman, V., Kostić, A., & Baayen, R. H. (2009). Paradigms bit by bit: An information-theoretic approach to the processing of paradigmatic structure in inflection and derivation. In J. P. Blevins, & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 214–252). Oxford, England: Oxford University Press.

Mirković, J., Seidenberg, M. S., & Joanisse, M.F. (2011). Rules vs. Statistics: Insights from a highly inflected language. *Cognitive Science, 35*, 638–681. https://doi.org/10.1111/j.1551–6709.2011.01174.x

Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition, 94*, 1–18. https://doi.org/10.1016/j.cognition.2003.10.015

Nenadić, F., Tucker, B.V., & Milin, P. (2016). *Relative entropy effects on lexical recognition of spoken Romanian verbs*. Paper presented at the conference "XXII naučni skup Empirijska istraživanja u psihologiji", Beograd, 18. – 20. March 2016. The Book of Abstracts, 52–53.

Plaut, D. C., & Gonnerman, L.M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes, 15*, 445–485.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005) Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica, 62*, 146–159. https://doi.org/10.1159/000090095

R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ramscar, M., Dye, M., & Klein, J. (2013) Children value informativity over logic in word learning. *Psychological Science, 24*(6), 1017–1023. https://doi.org/10.1177/0956797612460691

Ramscar, M., & Port, R. (2016) How spoken languages work in the absence of an inventory of discrete units. *Language Sciences, 53*(A), 58–74. https://doi.org/10.1016/j.langsci.2015.08.002

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science, 34*, 909 –957. https://doi.org/10.1111/j.1551–6709.2009.01092.x

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical Conditioning II* (pp. 64–99). Appleton-Century-Crofts.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Seidenberg, M. S., & Gonnerman, L.M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4,* 353–361. https://doi.org/10.1016/S1364–6613(00)01515–1

Seidenberg, M. S., & McClelland, J.L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523–568. https://doi.org/10.1037/0033–295X.96.4.523

Seidenberg, M. S., & Plaut, D.C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science, 38,* 1190–1228. https://doi.org/10.1111/cogs.12147

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *XXVII,* 379–423.

Stanojčić, Ž., & Popović, Lj., (1992). *Gramatika srpskog jezika* [Grammar of Serbian language]. Beograd: Zavod za udžbenike i nastavna sredstva.

Tabak, W., Schreuder, R., & Baayen, R. H. (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. In M. Reis, & S. Kepser (Eds.), *Linguistic Evidence* (pp. 529–555). Mouton. https://doi.org/10.1515/9783110197549.529

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition, 6*, 410–415. https://doi.org/10.3758/BF03197599

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9*(3), 271–294. https://doi.org/10.1080/01690969408402120

Taft, M., & Forster, K., (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638–647. https://doi.org/10.1016/S0022–5371(75)80051-X

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2015). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 1.0.1.https://cran.r-project.org/web/packages/itsadug/index.html

Wood, S. N. (2006) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

Wood, S. N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B), 73*(1), 3–36. https://doi.org/10.1111/j.1467–9868.2010.00749.x

Woollams, A. M., Joanisse, M. F., & Patterson, K. (2009). Past-tense generation from form and meaning: Behavioural data and simulation evidence. *Journal of Memory and Language*, *61*, 55–76. https://doi.org/10.1016/j.jml.2009.02.002

Wurm, L. H., Ernestus, M., Schreuder, R., & Baayen, R. H. (2006) Dynamics of the Auditory Comprehension of Prefixed Words: Cohort Entropies and Conditional Root Uniqueness Points.*The Mental Lexicon, 1*, 125–146. https://doi.org/10.1075/ml.1.1.08wur

Simultani efekti flektivnih paradigmi
i klasa u obradi srpskih glagola

Dušica Filipović Đurđević i Isidora Gatarić

Odsek za psihologiju, Filozofski fakultet, Univerzitet u Novom Sadu, Srbija

U ovom radu pokazujemo da na obradu flektivnih oblika glagola istovremeno utiču distribuciona svojstva njihove flektivne paradigme (svi flektivni oblici datog glagola) i njihove flektivne klase (svi glagoli koji se menjaju na isti način). Time na novu klasu reči uopštavamo nalaz koji je prethodno dobijen na imenicama. Pokazujemo da odstupanje distribucije frekvencija unutar flektivne paradigme od distribucije frekvencija unutar flektivne klase (operacionalizovano kao relativna entropija između dve distribucije frekvencija) ometa obradu. Predstavljamo rezultate eksperimenta sa zadatkom leksičke odluke i rezultate simulacije koja je sprovedena u Naivnom diskriminativnom čitaocu (Naive Discriminative Reader), jednostavnom računskom modelu zasnovanom na osnovnim principima učenja. Pokazujemo da je relativna entropija između flektivne paradigme i flektivne klase dobar prediktor kako opaženih, tako i simuliranih vremena obrade. Ovim doprinosimo korpusu istraživanja koja ispituju vezu između kognitivne obrade i opisa jezika zasnovanih na teoriji informacije. Dodatno, pokazujemo i da efekat relativne entropije na obradu morfologije može da se izrodi kao posledica principa diskriminativnog učenja u sistemu koji mapira ulazne znakove na ishode, a bez formulisanja morfologije same za sebe.

Ključne reči: flektivna morfologija, glagoli, paradigme i klase, relativna entropija, naivni diskriminativni čitalac (Naive Discriminative Reader)

Appendix A

Stimuli presented in the experiment: Lemma (illustrated by infinitive form), Lemma frequency, inflected forms presented to the participants (1st and 2nd person singular), their surface frequencies and relative entropy; all frequencies are obtained in a corpus of 2 million words.

Infinitive Form	Lemma frequency	1st person Singular	Surface frequency	2nd person singular	Surface frequency	Relative entropy
ZNATI	4123	ZNAM	1412	ZNAŠ	357	.09
IMATI	3905	IMAM	260	IMAŠ	84	.06
DATI	1953	DAM	85	DAŠ	15	.26
GLEDATI	1781	GLEDAM	426	GLEDAŠ	92	.04
ČEKATI	1258	ČEKAM	228	ČEKAŠ	44	.03
MORATI	1149	MORAM	138	MORAŠ	43	.06
PEVATI	1143	PEVAM	93	PEVAŠ	15	.04
PADATI	961	PADAM	37	PADAŠ	12	.14
IGRATI	777	IGRAM	24	IGRAŠ	6	.09
SPAVATI	731	SPAVAM	22	SPAVAŠ	40	.13
SLUŠATI	718	SLUŠAM	222	SLUŠAŠ	29	.11
PITATI	675	PITAM	167	PITAŠ	30	.06
SANJATI	626	SANJAM	139	SANJAŠ	28	.05
OSEĆATI	541	OSEĆAM	158	OSEĆAŠ	24	.12
MENJATI	429	MENJAM	174	MENJAŠ	1	.16
VRAĆATI	424	VRAĆAM	81	VRAĆAŠ	17	.05
PRIČATI	399	PRIČAM	34	PRIČAŠ	8	.05
SMATRATI	386	SMATRAM	15	SMATRAŠ	1	0
LUTATI	378	LUTAM	108	LUTAŠ	22	.08
ČUVATI	360	ČUVAM	39	ČUVAŠ	18	.06
PRUŽATI	330	PRUŽAM	66	PRUŽAŠ	16	.07
POGLEDATI	324	POGLEDAM	36	POGLEDAŠ	13	.26
PREDATI	281	PREDAM	18	PREDAŠ	4	.28
BACATI	275	BACAM	32	BACAŠ	9	.06
PRIMATI	270	PRIMAM	31	PRIMAŠ	6	.16
OTVARATI	267	OTVARAM	37	OTVARAŠ	4	.07
KORAČATI	263	KORAČAM	62	KORAČAŠ	12	.08
CVETATI	258	CVETAM	6	CVETAŠ	9	.13
STVARATI	257	STVARAM	5	STVARAŠ	0	.13
OSTAVLJATI	245	OSTAVLJAM	61	OSTAVLJAŠ	13	.11
ČITATI	239	ČITAM	39	ČITAŠ	15	.06
POZDRAVLJATI	225	POZDRAVLJAM	67	POZDRAVLJAŠ	2	.12
SPREMATI	219	SPREMAM	26	SPREMAŠ	10	.08
KUCATI	219	KUCAM	12	KUCAŠ	6	.14
PUCATI	217	PUCAM	8	PUCAŠ	2	.09
SKRIVATI	210	SKRIVAM	18	SKRIVAŠ	5	.09
DOČEKATI	210	DOČEKAM	8	DOČEKAŠ	1	.23

Infinitive Form	Lemma frequency	1st person Singular	Surface frequency	2nd person singular	Surface frequency	Relative entropy
PLAĆATI	210	PLAĆAM	58	PLAĆAŠ	2	.15
ODGOVARATI	204	ODGOVARAM	8	ODGOVARAŠ	5	.11
BLISTATI	201	BLISTAM	1	BLISTAŠ	1	.20
HODATI	200	HODAM	34	HODAŠ	12	.06
SPUŠTATI	198	SPUŠTAM	23	SPUŠTAŠ	7	.11
HVATATI	189	HVATAM	20	HVATAŠ	7	.09
DODATI	182	DODAM	3	DODAŠ	1	.46
STAVLJATI	176	STAVLJAM	24	STAVLJAŠ	1	.13
IZGLEDATI	171	IZGLEDAM	7	IZGLEDAŠ	3	.17
SNIVATI	169	SNIVAM	25	SNIVAŠ	5	.09
POZNATI	166	POZNAM	36	POZNAŠ	14	.27
ŠETATI	164	ŠETAM	10	ŠETAŠ	2	.09
VLADATI	161	VLADAM	4	VLADAŠ	7	.15
RAZVIJATI	149	RAZVIJAM	4	RAZVIJAŠ	1	.13
RAZGOVARATI	148	RAZGOVARAM	14	RAZGOVARAŠ	1	.11
KIDATI	147	KIDAM	19	KIDAŠ	4	.11
ZAHTEVATI	143	ZAHTEVAM	2	ZAHTEVAŠ	0	.14
UGLEDATI	142	UGLEDAM	24	UGLEDAŠ	4	.26
SKIDATI	134	SKIDAM	16	SKIDAŠ	5	.05
PRIPADATI	133	PRIPADAM	9	PRIPADAŠ	3	.11
POKUŠATI	130	POKUŠAM	0	POKUŠAŠ	1	.37
KOPATI	128	KOPAM	12	KOPAŠ	4	.07
UDARATI	127	UDARAM	7	UDARAŠ	1	.13
UPOZNATI	126	UPOZNAM	3	UPOZNAŠ	2	.19
UZIMATI	124	UZIMAM	13	UZIMAŠ	2	.08
DOZIVATI	120	DOZIVAM	23	DOZIVAŠ	4	.07
POSMATRATI	117	POSMATRAM	10	POSMATRAŠ	1	0
PROČITATI	116	PROČITAM	5	PROČITAŠ	1	.3
SAZNATI	112	SAZNAM	8	SAZNAŠ	6	.25
UBIJATI	112	UBIJAM	7	UBIJAŠ	0	.15
PRIZNATI	111	PRIZNAM	7	PRIZNAŠ	2	.23
PUŠTATI	104	PUŠTAM	20	PUŠTAŠ	5	0
NAPADATI	101	NAPADAM	0	NAPADAŠ	1	.23
BIRATI	94	BIRAM	8	BIRAŠ	0	.13
SIPATI	94	SIPAM	3	SIPAŠ	1	.15
IZDATI	91	IZDAM	1	IZDAŠ	0	.31
STUPATI	90	STUPAM	7	STUPAŠ	1	.09
REŠAVATI	89	REŠAVAM	1	REŠAVAŠ	1	.19
PROCVETATI	89	PROCVETAM	0	PROCVETAŠ	1	.26
SAČUVATI	87	SAČUVAM	5	SAČUVAŠ	2	.21
SAVLADATI	84	SAVLADAM	2	SAVLADAŠ	0	.22
SNEVATI	81	SNEVAM	11	SNEVAŠ	0	.09
ZAGLEDATI	79	ZAGLEDAM	12	ZAGLEDAŠ	2	.15
POKRIVATI	79	POKRIVAM	4	POKRIVAŠ	4	.17
ZIDATI	79	ZIDAM	9	ZIDAŠ	1	.13

Infinitive Form	Lemma frequency	1st person Singular	Surface frequency	2nd person singular	Surface frequency	Relative entropy
PRODATI	77	PRODAM	2	PRODAŠ	0	.25
UŽIVATI	75	UŽIVAM	5	UŽIVAŠ	2	.09
OTIMATI	70	OTIMAM	6	OTIMAŠ	1	.21
MEŠATI	68	MEŠAM	8	MEŠAŠ	0	.12
ZASIJATI	66	ZASIJAM	0	ZASIJAŠ	0	.41
OSIGURATI	65	OSIGURAM	2	OSIGURAŠ	0	.39
POJAČATI	63	POJAČAM	1	POJAČAŠ	0	.59
STIŠATI	63	STIŠAM	0	STIŠAŠ	0	.29
STISKATI	63	STISKAM	4	STISKAŠ	0	.29
POBOLJŠATI	57	POBOLJŠAM	0	POBOLJŠAŠ	11	.53
ISPRIČATI	53	ISPRIČAM	9	ISPRIČAŠ	2	.2
ISKOPATI	52	ISKOPAM	1	ISKOPAŠ	2	.5
ODATI	52	ODAM	3	ODAŠ	1	.12
SAČEKATI	45	SAČEKAM	5	SAČEKAŠ	0	.25
JURIŠATI	42	JURIŠAM	5	JURIŠAŠ	1	.12
IŠČUPATI	41	IŠČUPAM	0	IŠČUPAŠ	1	.49
ZAKOPATI	40	ZAKOPAM	3	ZAKOPAŠ	2	.64
VREĐATI	39	VREĐAM	2	VREĐAŠ	1	.21
SASLUŠATI	32	SASLUŠAM	1	SASLUŠAŠ	0	.72
UTKATI	29	UTKAM	2	UTKAŠ	0	.33
OTERATI	24	OTERAM	1	OTERAŠ	0	.38
OPEVATI	23	OPEVAM	2	OPEVAŠ	1	.42
POKIDATI	22	POKIDAM	1	POKIDAŠ	0	.64
TRAŽITI	1641	TRAŽIM	266	TRAŽIŠ	45	.07
NOSITI	1526	NOSIM	150	NOSIŠ	34	.08
GOVORITI	1250	GOVORIM	157	GOVORIŠ	28	.08
VODITI	916	VODIM	33	VODIŠ	13	.06
MISLITI	902	MISLIM	270	MISLIŠ	67	.19
VRATITI	851	VRATIM	92	VRATIŠ	22	.06
RADITI	807	RADIM	25	RADIŠ	12	.07
JAVITI	747	JAVIM	13	JAVIŠ	5	.27
DOLAZITI	744	DOLAZIM	39	DOLAZIŠ	13	.16
OTVORITI	713	OTVORIM	30	OTVORIŠ	3	.14
ČINITI	701	ČINIM	18	ČINIŠ	12	.19
LJUBITI	593	LJUBIM	108	LJUBIŠ	27	.08
ZABORAVITI	570	ZABORAVIM	49	ZABORAVIŠ	22	.07
PROLAZITI	518	PROLAZIM	57	PROLAZIŠ	8	.16
UČINITI	496	UČINIM	15	UČINIŠ	4	.08
IZGUBITI	493	IZGUBIM	16	IZGUBIŠ	5	.18
OSTAVITI	492	OSTAVIM	17	OSTAVIŠ	4	0
MOLITI	492	MOLIM	183	MOLIŠ	10	.3
PRATIITI	462	PRIMIM	7	PRIMIŠ	3	0
RODITI	440	RODIM	12	RODIŠ	3	.13
PRUŽITI	413	PRUŽIM	26	PRUŽIŠ	5	.05
ZAVRŠITI	411	ZAVRŠIM	3	ZAVRŠIŠ	1	.39

Infinitive Form	Lemma frequency	1st person Singular	Surface frequency	2nd person singular	Surface frequency	Relative entropy
VRŠITI	410	VRŠIM	7	VRŠIŠ	0	.12
SLUTITI	400	SLUTIM	115	SLUTIŠ	35	.2
ODLAZITI	398	ODLAZIM	61	ODLAZIŠ	25	.13
GAZITI	395	GAZIM	60	GAZIŠ	13	.12
PLOVITI	383	PLOVIM	30	PLOVIŠ	3	.17
IZVRŠITI	371	IZVRŠIM	3	IZVRŠIŠ	0	.21
PUSTITI	365	PUSTIM	23	PUSTIŠ	10	.08
PRATITI	364	PRATIM	20	PRATIŠ	6	.06
SPUSTITI	325	SPUSTIM	24	SPUSTIŠ	6	.07
ZVONITI	322	ZVONIM	10	ZVONIŠ	1	.2
BACITI	313	BACIM	20	BACIŠ	7	.06
STVORITI	299	STVORIM	3	STVORIŠ	4	.09
PALITI	298	PALIM	24	PALIŠ	8	.12
UPUTITI	297	UPUTIM	9	UPUTIŠ	3	.28
ŠIRITI	292	ŠIRIM	21	ŠIRIŠ	10	.13
GRLITI	287	GRLIM	41	GRLIŠ	10	.11
LOMITI	281	LOMIM	18	LOMIŠ	3	.13
POBEDITI	280	POBEDIM	3	POBEDIŠ	1	.37
OSETITI	277	OSETIM	24	OSETIŠ	13	.19
NASTAVITI	277	NASTAVIM	6	NASTAVIŠ	1	.15
GUBITI	271	GUBIM	29	GUBIŠ	12	.11
STAVITI	270	STAVIM	11	STAVIŠ	3	.08
BRANITI	260	BRANIM	17	BRANIŠ	3	.12
PROBUDITI	258	PROBUDIM	16	PROBUDIŠ	8	.06
POSTAVITI	255	POSTAVIM	3	POSTAVIŠ	1	.25
REŠITI	245	REŠIM	4	REŠIŠ	0	.21
DONOSITI	241	DONOSIM	6	DONOSIŠ	10	.11
RUŠITI	241	RUŠIM	13	RUŠIŠ	4	.16
SHVATITI	239	SHVATIM	18	SHVATIŠ	5	.12
ODGOVORITI	234	ODGOVORIM	6	ODGOVORIŠ	3	.19
GRADITI	231	GRADIM	14	GRADIŠ	5	.11
OBJAVITI	224	OBJAVIM	0	OBJAVIŠ	0	.3
SLUŽITI	218	SLUŽIM	5	SLUŽIŠ	0	.07
DELITI	214	DELIM	11	DELIŠ	5	0
MUČITI	199	MUČIM	8	MUČIŠ	3	0
ŽALITI	198	ŽALIM	41	ŽALIŠ	12	.12
POSETITI	195	POSETIM	4	POSETIŠ	0	.26
OBEZBEDITI	192	OBEZBEDIM	0	OBEZBEDIŠ	0	.31
UČITI	190	UČIM	13	UČIŠ	4	.05
PRIHVATITI	183	PRIHVATIM	3	PRIHVATIŠ	0	.15
KUPITI	179	KUPIM	12	KUPIŠ	1	.11
KORISTITI	175	KORISTIM	2	KORISTIŠ	0	.15
OSTVARITI	174	OSTVARIM	1	OSTVARIŠ	0	.15
ZATVORITI	172	ZATVORIM	8	ZATVORIŠ	6	.17
ISPUNITI	170	ISPUNIM	1	ISPUNIŠ	0	.14

Infinitive Form	Lemma frequency	1st person Singular	Surface frequency	2nd person singular	Surface frequency	Relative entropy
OSLOBODITI	166	OSLOBODIM	6	OSLOBODIŠ	1	.14
NAPUSTITI	164	NAPUSTIM	1	NAPUSTIŠ	0	.16
UHVATITI	162	UHVATIM	11	UHVATIŠ	6	.04
PRETVORITI	162	PRETVORIM	4	PRETVORIŠ	2	.08
PRAVITI	154	PRAVIM	8	PRAVIŠ	6	.07
KRUŽITI	152	KRUŽIM	3	KRUŽIŠ	2	.23
ODREDITI	150	ODREDIM	1	ODREDIŠ	0	.31
NUDITI	150	NUDIM	11	NUDIŠ	4	.14
UDARITI	149	UDARIM	7	UDARIŠ	1	.09
NAUČITI	148	NAUČIM	2	NAUČIŠ	1	.2
PRETITI	146	PRETIM	1	PRETIŠ	3	.17
PREDLOŽITI	144	PREDLOŽIM	0	PREDLOŽIŠ	0	.39
OBAVITI	137	OBAVIM	1	OBAVIŠ	0	.36
HRANITI	136	HRANIM	12	HRANIŠ	4	.05
IZRAZITI	136	IZRAZIM	4	IZRAZIŠ	0	.3
POKLONITI	134	POKLONIM	9	POKLONIŠ	4	.11
SKLOPITI	134	SKLOPIM	12	SKLOPIŠ	3	.07
ZAGRLITI	131	ZAGRLIM	15	ZAGRLIŠ	5	0
SLAVITI	127	SLAVIM	9	SLAVIŠ	1	0
POMISLITI	126	POMISLIM	25	POMISLIŠ	9	.21
POSVETITI	124	POSVETIM	2	POSVETIŠ	0	.3
STREPITI	124	STREPIM	33	STREPIŠ	9	.25
UGASITI	120	UGASIM	3	UGASIŠ	0	.14
SRUŠITI	119	SRUŠIM	7	SRUŠIŠ	1	.09
NAPRAVITI	118	NAPRAVIM	6	NAPRAVIŠ	0	.12
PRIMETITI	116	PRIMETIM	1	PRIMETIŠ	0	.25
SAHRANITI	113	SAHRANIM	4	SAHRANIŠ	1	.58
SMIRITI	113	SMIRIM	16	SMIRIŠ	0	.14
ISKORISTITI	113	ISKORISTIM	2	ISKORISTIŠ	0	.19
POLJUBITI	110	POLJUBIM	11	POLJUBIŠ	4	.13
UNIŠTITI	109	UNIŠTIM	1	UNIŠTIŠ	0	.28
ZAPALITI	108	ZAPALIM	9	ZAPALIŠ	1	.11
ZAKLJUČITI	106	ZAKLJUČIM	1	ZAKLJUČIŠ	0	.27
KLIZITI	100	KLIZIM	2	KLIZIŠ	2	.22
IZMENITI	99	IZMENIM	0	IZMENIŠ	0	.24
PLATITI	99	PLATIM	4	PLATIŠ	1	.14
IZGRADITI	99	IZGRADIM	0	IZGRADIŠ	1	.27
PODELITI	97	PODELIM	3	PODELIŠ	2	.24
MERITI	97	MERIM	13	MERIŠ	2	.21
SPREČITI	95	SPREČIM	0	SPREČIŠ	1	.18
PROMENITI	94	PROMENIM	1	PROMENIŠ	1	.13
SLOMITI	93	SLOMIM	1	SLOMIŠ	0	.25
TEŠITI	92	TEŠIM	9	TEŠIŠ	1	.17
ZAMENITI	90	ZAMENIM	1	ZAMENIŠ	0	.11
PRIPREMITI	89	PRIPREMIM	0	PRIPREMIŠ	0	.16

Infinitive Form	Lemma frequency	1st person Singular	Surface frequency	2nd person singular	Surface frequency	Relative entropy
SKOČITI	89	SKOČIM	3	SKOČIŠ	0	.21
MOTRITI	87	MOTRIM	14	MOTRIŠ	1	.15
SPAZITI	85	SPAZIM	12	SPAZIŠ	3	.2
POTVRDITI	84	POTVRDIM	3	POTVRDIŠ	1	.19
SMANJITI	83	SMANJIM	1	SMANJIŠ	0	.14
PRODUŽITI	81	PRODUŽIM	9	PRODUŽIŠ	2	.11
SKLONITI	80	SKLONIM	10	SKLONIŠ	2	.13
UPOTREBITI	78	UPOTREBIM	1	UPOTREBIŠ	0	.19
SPREMITI	77	SPREMIM	4	SPREMIŠ	0	.12
OBORITI	73	OBORIM	2	OBORIŠ	0	.14
PROGOVORITI	71	PROGOVORIM	2	PROGOVORIŠ	2	.11
SKUPITI	71	SKUPIM	5	SKUPIŠ	0	.16
SLEDITI	69	SLEDIM	1	SLEDIŠ	1	.13
OBAVESTITI	68	OBAVESTIM	2	OBAVESTIŠ	1	.26
RAŠIRITI	68	RAŠIRIM	4	RAŠIRIŠ	0	.07
USPOSTAVITI	67	USPOSTAVIM	0	USPOSTAVIŠ	1	.15
ODBACITI	67	ODBACIM	0	ODBACIŠ	2	.23
PROSLAVITI	66	PROSLAVIM	2	PROSLAVIŠ	0	.29
UPALITI	66	UPALIM	3	UPALIŠ	2	.2
IZVADITI	65	IZVADIM	2	IZVADIŠ	2	.12
SASTAVITI	63	SASTAVIM	0	SASTAVIŠ	1	.3
PREBACITI	62	PREBACIM	0	PREBACIŠ	1	.13
POVERITI	60	POVERIM	5	POVERIŠ	0	.16
ZGAZITI	60	ZGAZIM	3	ZGAZIŠ	2	.08
SAGRADITI	59	SAGRADIM	1	SAGRADIŠ	0	.27
POGODITI	59	POGODIM	1	POGODIŠ	1	.2
ZAROBITI	58	ZAROBIM	1	ZAROBIŠ	0	.45
POPRAVITI	55	POPRAVIM	2	POPRAVIŠ	0	.2
ZARONITI	55	ZARONIM	7	ZARONIŠ	3	.18
SPALITI	53	SPALIM	3	SPALIŠ	2	.2
CEDITI	52	CEDIM	1	CEDIŠ	1	.12
PROPUSTITI	50	PROPUSTIM	6	PROPUSTIŠ	0	.18
OBESITI	44	OBESIM	1	OBESIŠ	1	.21
SMESTITI	39	SMESTIM	2	SMESTIŠ	1	.55
DOVRŠITI	38	DOVRŠIM	1	DOVRŠIŠ	0	.24
OSVETLITI	37	OSVETLIM	0	OSVETLIŠ	0	.26
UKLJUČITI	37	UKLJUČIM	0	UKLJUČIŠ	0	.5
ISPRATITI	30	ISPRATIM	1	ISPRATIŠ	0	.33
IZMERITI	27	IZMERIM	1	IZMERIŠ	0	.22
ZASADITI	26	ZASADIM	1	ZASADIŠ	0	.4

Appendix B

Pair-wise Pearson correlation coefficients of predictors (all are significant at the level of $p < .01$)

	Word length in letters	Lemma frequency	Surface frequency
Word length in letters			
Lemma frequency	-.39		
Surface frequency	-.38	.66	
Relative entropy	.26	-.48	-.47