Article:

Filipović Đurđević, D. and Milin, P. orcid.org/0000-0001-9708-7031 (2019) Information and learning in processing adjective inflection. Cortex, 116. pp. 209-227. ISSN 0010-9452

https://doi.org/10.1016/j.cortex.2018.07.020

Information and Learning in Processing of Adjective Inflection

Dušica Filipović Đurđević[*a] and Petar Milin[b]

[a]Department of Psychology, Faculty of Philosophy, University of Novi Sad, Dr Zorana Đinđića 2, Novi Sad, Serbia
[b]Department of Journalism Studies, University of Sheffield, Sheffield, United Kingdom

* Correspondence[1]:
Dr Dušica Filipović Đurđević
Department of Psychology
Faculty of Philosophy
Dr Zorana Đinđića 2
21000 Novi Sad
Serbia
tel/fax: +381 21 458 948
E-mail: dfdurdevic@ff.uns.ac.rs

[1] Dušica Filipović Đurđević has moved and currently holds a position at the Department of Psychology, Faculty of Philosophy, University of Belgrade, Čika Ljubina 18-20, 11000 Beograd, Serbia, Tel: +381 11 2630 542, Fax: +381 11 2639 356, E-mail: dusica.djurdjevic@f.bg.ac.rs

Abstract

We investigated the processing of inflected Serbian adjective forms to bring together quantitative linguistic measures from two frameworks – information theory and discrimination learning. From each framework we derived several quantitative descriptions of an inflectional morphological system and fitted two separate regression models to the processing latencies that were elicited by inflected adjectival forms presented in a visual lexical decision task. The model, which was based on lexical distributional and information theory revealed a dynamic interplay of information. The information was sensitive to syntagmatic and paradigmatic dimensions of variation; the paradigmatic information (formalized as respective relative entropies) was also modulated by lemma frequency. The discrimination learning based model revealed an equally complex pattern, involving several learning-based variables. The two models revealed strikingly similar patterns of results, as confirmed by the very high proportion of shared variance in model predictions (85.83%). Our findings add to the body of research demonstrating that complex morphological phenomena can arise as a consequence of the basic principles of discrimination learning. Learning discriminatively about inflectional paradigms and classes, and about their contextual or syntagmatic embedding, sheds light on human language-processing efficiency and on the fascinating complexity of naturally emerged language systems.

Keywords: adjectives, discrimination learning, inflectional morphology, prepositional phrases, relative entropy

1. Introduction

In this paper we take advantage of the rich inflectional morphology of the Serbian language to investigate further how the human cognitive system engages with probabilistic features of language.[2] In this study we make use of probabilistic formalizations originally developed as information-theoretic measures (cf., Shannon, 1948; MacKay, 2003). We then relate this cognitive *sensitivity to probabilistic patterns* to the fundamental processes of learning and discrimination learning in particular, as inspired by Rescorla and Wagner (1972) and established by Ramscar and associates (Ramscar & Yarlett, 2007; Ramscar, Yarlett, Dye, Denny, & Thorpe 2010; Arnon & Ramscar, 2012). We aim to demonstrate that sensitivity to patterns, which are identified via information-theoric tools, can arise straightforwardly from principles of discrimination learning.

1.1. The information-theoretic framework

Information theory (Shannon, 1948) has been shown to be a fruitful framework for describing various phenomena in language and related to language. Several quantitative descriptions of language based on information theory have been established as predictors of response latencies (Frank, 2013; Hale, 2001; 2003; 2006; Kostić, 1991; 1995; Moscoso del Prado Martín, Kostić, & Baayen, 2004; Schmidtke, Kuperman, Gagné, & Spalding, 2016), eye-fixations (Boston, Hale, Patil, Kliegl, & Vasishth, 2008; Demberg, & Keller, 2008), as well as neural activity during word processing (Frank, Otten, Galli, & Vigliocco, 2015; Henderson, Choi, Lowder, & Ferreira, 2016; Hendrix & Baayen, 2014; Linzen, Marantz, & Pylkkänen, 2013). In this paper, we focus on paradigmatic relative entropy, whose effects were originally demonstrated in the processing of inflected nouns (Milin, Filipović Đurđević, & Moscoso del Prado Martin, 2009).

Inflection is used to denote the specific syntactic role of a given word. For example, the Serbian feminine noun form *vila* (*fairy*; Nom. Sg.) could indicate the subject in a sentence (e.g. *Vila je nestala*. /*The fairy disappeared*./), whereas *vilu* (*fairy*; Acc. Sg.) could function as (among other roles) an object of some action (e.g. *Video je vilu*. /*He saw a fairy*/). The set of all possible inflected forms of a given lemma (word stem) constitutes its *inflectional paradigm*, which can be described probabilistically in terms of the occurrence counts (i.e., frequencies) of respective inflected forms. At the same time, each lemma can also be attributed to a set consisting of all words that inflect in exactly the same way – i.e., attaching the same affixes to the stem to form possible inflected variants. This constitutes an *inflection class* (e.g., regular feminine nouns). Inflection classes can be described in terms of occurrences, as in inflectional paradigms, where we must take the cumulative frequencies of all the class members appearing in a given inflected form (i.e., exhibiting a particular affix).

For each individual lemma (or word stem), the shapes of the two frequency distributions – at the level of an individual paradigm and at the level of the whole class – can be compared. It turns out that some lemmata have a paradigm-based distribution that is very similar to the corresponding class-based distribution, whereas other words have a paradigm-based distribution that is very different from the distribution of their class. To put it simply, lemmata can be positioned along a continuum with respect to the degree of difference between two frequency distributions, local (paradigm) and global (class). A standard measure

[2] For a comprehensive overview of statistical learning and language, see Jost & Christiansen, 2016; Armstrong, Frost, & Christiansen, 2017.

from Information Theory known as the relative entropy or Kullback-Leibler divergence ($D(p||q)$; Cover & Thomas, 1991), can be used as an indicator of the dissimilarity between two probability distributions: the more two distributions diverge, the larger the relative entropy. Milin et al. (2009) were the first to demonstrate that the divergence of paradigm-based probability distribution from class-based probability distribution is predictive of word processing latency and accuracy, even when words are presented in isolation. Their finding revealed that the time spent in recognizing an inflected noun form is influenced by the difference between two probability distributions (which are obtained as estimates from frequency counts): the greater the divergence, the longer the processing.

The findings of Milin et al. (2009) added to an existing body of research by demonstrating that local and global form-related features (i.e., morphological features) affected processing of 'morphologically complex' words.[3] Baayen, Milin, Filipović Đurđević, Hendrix, and Marreli (2011) then extended this finding to processing of Serbian inflected nouns in the context of a sentence, and/or a prime, and applied the same approach to English prepositional phrases. The effects demonstrated in the declensional domain also arise in verb paradigms and classes (conjugations), as shown by Filipović Đurđević and Gatarić (in press) for visually presented Serbian verbs, and by Nenadić, Tucker, and Milin (2016) for auditorily-presented Romanian verbs. Additional supporting evidence has been reported in the processing of English words (Kuperman et al., 2010; Milin, Kuperman et al., 2009). This behavioural evidence has also been reinforced by studies that have shown that electrophysiological responses in the brain can be predicted from relative entropy (Hendrix & Baayen, 2014; Linzen, Marantz, & Pylkkänen, 2013).

1.2. The discrimination learning framework

Baayen et al. (2011) interpreted the effects of paradigmatic relative entropy by proposing a model based on Naïve Discrimination Learning (NDL). The model is based on the set of equations as introduced by Rescorla and Wagner (1972), which are themselves related to the learning rule of Widrow and Hoff (1960), who proposed incremental updating to minimise prediction error (essentially, iteratively obtained regression-like weights). Rescorla-Wagner discrimination learning was firstly used as an explanatory framework for language acquisition and processing by Ellis (2006) and Ramscar and colleagues (Ramscar & Yarlett, 2007; Ramscar et al., 2010; Ramscar & Port, 2016).

The NDL model is built on a simple two-layer network architecture, in which one layer is devoted to the input stimulation, and the other one to the output. In the original formulation (Baayen et al., 2011), the input was represented by bigraphs that constituted the word form, and the output captured lexical and grammatical units. In later modifications of the model, the output was devoted to so-called 'lexomes' (a term chosen to avoid any baggage associated with related notions such as 'lemma' or 'lexeme'). Lexomes were defined as pointers to 'meanings' which are realized contextually, distributed in a high-dimensional semantic vector space (for details consult Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017). Essentially, the model learns to discriminate cues that are good predictors of an outcome from those that are non-discriminative, where such 'discriminability' (i.e., predictivity) is indicated as the learning weights – i.e., association strengths among cues and outcomes. The weights' updating is based on the equations of Rescorla and Wagner (1972): the weights remain the same in the absence of the given cue, their value increases if both cue

[3] See Milin, Kuperman, Kostić, & Baayen, 2009 for a more detailed and formal account of information-theoretic approaches.

and outcome are present, and decreases if the cue is present in the absence of the outcome (for technical details consult Baayen et al., 2011; Milin et al., 2017). Again, the weights serve as indicators of predictive (or discriminative) power of the given cues for the given outcomes.

Importantly for the present study, it was shown that this model captures the effect of relative entropy between a word's paradigm and class (Baayen et al., 2011; Filipović Đurđević & Gatarić, in press), as well as numerous other set-related effects, such as the effect of morphological family size (De Jong, Schreuder, & Baayen, 2000), the effect of inflectional entropy (Moscoso del Prado Martín, Kostić, & Baayen, 2004), and so on. The authors argued that the observed paradigmatic effects arise as a consequence of the learned discrimination of cues for grammatical case outcomes, where discrimination weights indicate the predictive potential of a given bigram input cue for the case outcome. In particular, the observed paradigmatic effects arise from dynamic competition among input cues when discriminating paradigmatic relatives (i.e., variants of the same lemma).

The NDL model introduced in Baayen et al. (2011) has evolved during the course of years (Baayen, Milin, & Ramscar, 2016; Milin et al., 2017). In addition to some other differences (which will be discussed in section 1.3.2), this model permitted the derivation of several quantitative measures from the matrix (i.e., network) of learned discrimination weights, which showed compelling predictive potential across language processing tasks. Although Milin et al. (2017) demonstrated the predictive validity of these measures, to the best of best knowledge, these new measures have never been explicitly linked to measures derived from an information-theoretic framework, nor have the two categories of measures (information theory based, and discrimination learning based) been compared, or tested against each other.

1.3. Current goal

In this paper, we examine inflected Serbian adjectives, guided by four specific research goals. First, we aim to demonstrate that the information-theoretic measure of relative entropy can predict the recognition time of inflected adjectival forms, following the earlier results of Milin et al. (2009) and others. We also probe simultaneously paradigmatic (Milin et al., 2009) and syntagmatic relative entropy effects (similar to Baayen et al., 2011). Second, we test several measures derived from the Naïve Discrimination Learning framework (Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017) as predictors of the recognition latencies of the same set of inflected adjectival forms. Finally, we aim to compare the two sets of measures (the one based on the information-theoric framework, and the other based on a Naïve Discrimination Learning framework), and demonstrate that the two sets of measures produce 'mirroring' effects on processing latencies. Additionally, we show that discrimination learning based measures show several advantages compared to measures which are based on information theory.

The adjectives in Serbian represent a very fruitful ground for testing various hypotheses regarding the processing of grammatical features. They are defined as the non-autonomous, or dependent PoS category, as they are typically coupled with a noun (Jakić, 2016; Stanojčić & Popović, 1992). As in other languages, they carry semantic information and help to identify the object denoted by the noun. They are also highly inflected and help disambiguate the grammatical case, number, and gender of the noun. However, grammatical features are not unambiguously mapped onto inflected forms (e.g. a single inflected form can point to multiple case/number combinations and multiple genders simultaneously). Additionally, unlike nouns, which have intrinsic gender, adjectives can appear in all gender categories. To fully disambiguate all grammatical features, it is necessary to consider the full adjective-noun pair, and sometimes a triplet consisting of the preposition, adjective and noun. Bearing in mind that these constituents appear in sequences (consecutively), with each carrying partial disambiguating information, they form an ideal ground for studying the sensitivity of the human processing system to rich language patterns of sequential dependencies. In studies that focus on the processing of isolated words, inflected adjectival forms are the hardest to disambiguate, thus they represent the ideal candidate for our current inquiry.

1.3.1. Lexical-distributional and information-theoretic predictors

Whereas Milin et al. (2009) focused only on paradigmatic entropy (the relative entropy derived from distributions of individual Serbian inflected forms), and Baayen et al. (2011) investigated syntagmatic entropy (the relative entropy derived from frequencies of English prepositional phrases), here we will simultaneously investigate both variants of relative entropy calculated for the same set of words of a highly inflected language. We will take a set of Serbian adjectives and calculate both paradigmatic and syntagmatic relative entropy.

1.3.1.1. Paradigmatic relative entropy

Previous research on inflected Serbian adjectives revealed that the appropriate adjectival paradigm/class is the one that includes frequencies of grammatical case/number combinations regardless of grammatical gender (Filipović Đurđević & Kostić 2003; 2004). Therefore, we calculated paradigmatic relative entropy based on this specification of adjectival paradigms and classes, by applying equation 1, given in Table 1.

$$\begin{aligned} D(p(x)||q(x)) &= \sum_{x} p(x) log \frac{p(x)}{q(x)} \\ &= \sum_{i=1}^{n} f(w_i)/f(w) log \frac{f(w_i)/f(w)}{f(e_i)/f(e)} \qquad (1) \end{aligned}$$

Equation (1) is taken from Milin et al. (2009) and adapted to fit the specifications of adjectival inflection. Here, *p(x)* denotes the frequency distribution of grammatical case/number combinations of the inflectional paradigm for the given adjective, and *q(x)* denotes the distribution of cumulative frequencies of grammatical case/number combinations for all adjectives, i.e. adjectival inflectional class. The subscript *i* denotes unique case/number combinations within a given distribution. The expression $f(w_i)$ represents the frequency of the *i*-th case/number category of the adjective *w*, and *f(w)* is the cumulative frequency of all the case/number combinations of the adjective *w*, that is, stem frequency of *w*. Likewise, $f(e_i)$ denotes the frequency of the *i*-th case/number category within the whole adjectival inflectional class, that is, the sum of frequencies of *i*-th case/number category of all the adjectives. Finally, *f(e)* stands for the cumulative frequency of all case/number categories, for all adjectives, i.e. for all the case/number categories across the whole inflectional class.

Table 1. Frequencies of unique combinations of case and number (merged across gender categories) for the adjective *nov* (new), their cumulative analogues from the corresponding inflectional class, and the result of applying Equation (1) to those frequencies.

Case (paradigm)	$f(w_i)$	$p(i)=f(w_i)/f(w)$	Case (class)	$f(e_i)$	$q(i)=f(e_i)/f(e)$	$p(i)log(p(i)/q(i))$
Nom. Sg.	813	0.25	*Nom. Sg.*	60101	0.30	-0.02
Gen. Sg.	391	0.12	*Gen. Sg.*	26970	0.13	0
Dat. Sg.	51	0.02	*Dat. Sg.*	3023	0.02	0
Acc. Sg.	585	0.18	*Acc. Sg.*	27887	0.14	0.02
Ins. Sg.	149	0.05	*Ins. Sg.*	8855	0.04	0
Loc. Sg.	198	0.06	*Loc. Sg.*	12610	0.06	0
Nom. Pl.	312	0.1	*Nom. Pl.*	24461	0.12	-0.01
Gen. Pl.	299	0.09	*Gen. Pl.*	19567	0.10	0
Dat. Pl.	24	0.01	*Dat. Pl.*	1148	0.01	0
Acc. Pl.	378	0.12	*Acc. Pl.*	8298	0.04	0.06
Ins. Pl.	49	0.01	*Ins. Pl.*	3858	0.02	0
Loc. Pl.	32	0.01	*Loc. Pl.*	4503	0.02	0
	$f(w)$=3281			$f(e)$=201281		$D(p\|\|q)$=0.05

1.3.1.2.Syntagmatic relative entropy

A full grasp of the *informativity* of adjectives (and other PoS) can only be understood contextually: their *syntagmatic uncertainty* is realized over time-distributed units such as words or phrases and sentences, as they naturally occur in a typical discourse. Conversely, *paradigmatic uncertainty* arises at each point in time and within a set of the given word's forms. We could also say that communication – as exchange of information by 'economizing' uncertainty – is realized simultaneously – *longitudinally and latitudinally* – to achieve maximum efficiency. If this is so, then we cannot focus on one dimension of uncertainty and ignore the other. In our testbed language, Serbian, an inflected form of an adjective *poznat-og* (familiar) can denote both genitive and accusative case, but pre-adjectival prepositions *od* (from) or *na* (on) resolve the apparent case ambiguity completely (genitive: *od poznatog* / from the familiar; accusative: *na poznatog* / on the familiar).

We therefore develop our previous hypothesis further by assuming that the full span of adjectival informativity must be realized in an interplay of the preposition and inflected adjectival form. This explicitly takes into account an adjective's syntagmatic and paradigmatic (or longitudinal and latitudinal) uncertainty. More formally, we operationalize adjectival paradigms as the sets of pairs of co-occurring prepositions and inflected variants of the given adjective. Similarly, the global adjectival class is operationalized as the set of co-occurring prepositions and inflected variants of all adjectives (a preposition and any adjective with the given suffix) combined. In this respect, we are taking the approach proposed by Baayen and colleagues (2011) who demonstrated phrasal paradigmatic effects in processing of English nouns[4] and consolidating it with the pioneering work of Milin et al. (2009) to provide a unified measure of the syntagmatic and paradigmatic complexity of adjectival processing.

The application of phrasal frequencies of preposition-article-noun trigrams (Baayen, et al., 2011) was a way to demonstrate paradigmatic effects in a language that does not manifest rich inflectional morphology on nouns. In order to take advantage of the rich Serbian morphology, we introduced a slight modification of Equation (1). Specifically, we

[4] Hendrix & Baayen 2014 demonstrated the same effect by recording EEG signals in a speech production task.

introduced joint probabilities of prepositions and inflected adjectival variants (Cover & Thomas, 1991):

$$D(p(x,y)||q(x,y)) = \sum_{x} \sum_{y} p(x,y) log \frac{p(x,y)}{q(x,y)}$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{m} f(w_i, r_j)/f(w) log \frac{f(w_i, r_j)/f(w)}{f(e_i, r_j)/f(e)} \quad (2)$$

In equation (2), *p(x,y)* refers to the joint probability of the given preposition and the given inflected adjectival variant – i.e., the probability of a specific preposition-adjective phrase (e.g. *od poznatog* / from the familiar), as determined by the corpus. Thus, *p(x,y)* mirrors the probability distribution of the paradigm and is the analogue of *p(x)* in Equation (1). At the same time, *q(x,y)* denotes the joint probability of the given preposition and the given inflected variant of all adjectives combined (i.e., summed). This is the probability of the given preposition appearing before any adjective ending with the given inflectional suffix (e.g. *od* ____*–og*). Hence, *q(x,y)* mirrors the probability distribution of the class and is the analogue to *q(x)* in Equation 1. The counter *i=1, ...n* refers to the number of inflected variants of an adjective, and the counter *j=1, ...m* marks the number of prepositions in the sample. Analogously, $f(w_i, r_j)$ denotes the phrasal frequency of a preposition-adjectival pair (a particular preposition and a particular inflected form of the given adjective) and *f(w)* marks the lemma frequency of the adjective. Finally, $f(e_i, r_j)$ denotes the cumulative phrasal frequency of the given preposition and all adjectives ending with the given inflectional exponent, and *f(e)* refers to cumulative frequency of all the adjectives in the sample.

As this equation is being introduced for the first time, we will illustrate its application step by step. Table 2 and Equation 3 illustrate how (2) can be applied to calculate the relative entropy of the adjective *poznat* (familiar). In order to do so, we must first obtain frequencies of each preposition that precedes each inflected form of this particular adjective (e.g. *f(u poznat)*=161; *f(u poznatog)*=13, … *f(iz poznato)*=27), etc.). Next, we transform these frequencies to probabilities by dividing them by the lemma frequency of the same adjective (e.g. *p(u poznat)*=161/1774.4=0.09073). Then we obtain frequencies of each preposition preceding any adjective ending with a given inflectional suffix (e.g. *f(u* ___ *- ø)*=1245.8; *f(u* ___*-og)=172*, … *f(u* ___ *-o)=297.4*) and transform them to probabilities by dividing each frequency with the summed frequency of all the adjectives in the sample (e.g. *p(u* ___*-ø)=1245.8/103932.7=0.01199*). Finally, we apply (2) to calculate the syntagmatic relative entropy of the adjective *poznat*:

$$D(p(preposition, poznat)||q(preposition, adjective)) = \sum_{x} \sum_{y} p(x,y) log \frac{p(x,y)}{q(x,y)}$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{m} f(w_i, r_j)/f(w) log \frac{f(w_i, r_j)/f(w)}{f(e_i, r_j)/f(e)}$$
$$= \left[161/1774.4 \times \log_{10} \frac{161/1774.4}{1245.8/103932.7}\right] + \cdots$$

$$\ldots + \left[27/1774.4 \times \log_{10} \frac{27/1774.4}{297.4/103932.7}\right]$$
$$= \left[0.09073 \times \log_{10} \frac{0.09073}{0.01199}\right] + \cdots$$
$$\ldots + \left[0.01522 \times \log_{10} \frac{0.01522}{0.00286}\right]$$
$$= 0.07975 + \cdots + 0.01105$$
$$= 0.219 \quad (3)$$

This approach provides more information than the method applied in Baayen et al. (2011). It implies both the disambiguation of the adjectival form (as described previously) and of the preposition (some prepositions can denote multiple cases, depending on the inflected form that they co-occur with; e.g. *u sobi* [in the room] denotes locative, whereas *u sobu* [into the room] denotes accusative).

Table 2. Preposition-adjective phrasal frequencies – frequencies of selected prepositions immediately preceding the given infected form of the given adjective (paradigm) and immediately preceding the given inflected form of any adjective (class) and the process of applying Equation (2) to those frequencies.

		j=1: u (in)	***j=2: na (on)***	***j=3: za (for)***	***j=4: od (from)***	***j=5: iz (from)***	
		$f(w_ir_j)$					
i=1	*poznat- ø*	161	33	18	25	26	
i=2	*poznat-og*	13	14	52	32	21	
i=3	*poznat-om*	111	38	1	0.1	6	
i=4	*poznat-im*	55	37	2	0.1	2	
i=5	*poznat-i*	88	29	34	12	8	
i=6	*poznat-ih*	21	6	2	114	66	
i=7	*poznat-e*	48	30	32	19	42	
i=8	*poznat-a*	86	16	23	10	16	
i=9	*poznat-oj*	82	24	0.1	0.1	2	
i=10	*poznat-u*	15	27	20	1	3	
i=11	*poznat-o*	140	41	26	17	27	*f(w)*=1774.4
		$p(ij)=f(w_ir_j)/f(w)$					
i=1	*poznat- ø*	0.09073	0.0186	0.01014	0.01409	0.01465	
i=2	*poznat-og*	0.00733	0.00789	0.02931	0.01803	0.01183	
i=3	*poznat-om*	0.06256	0.02142	0.00056	0.00006	0.00338	
i=4	*poznat-im*	0.031	0.02085	0.00113	0.00006	0.00113	
i=5	*poznat-i*	0.04959	0.01634	0.01916	0.00676	0.00451	
i=6	*poznat-ih*	0.01183	0.00338	0.00113	0.06425	0.0372	
i=7	*poznat-e*	0.02705	0.01691	0.01803	0.01071	0.02367	
i=8	*poznat-a*	0.04847	0.00902	0.01296	0.00564	0.00902	
i=9	*poznat-oj*	0.04621	0.01353	0.00006	0.00006	0.00113	
i=10	*poznat-u*	0.00845	0.01522	0.01127	0.00056	0.00169	
i=11	*poznat-o*	0.0789	0.02311	0.01465	0.00958	0.01522	

		$f(e_i r_j)$					
i=1	*- ø*	1245.8	1429.3	1241.5	219	69.1	
i=2	*-og*	172	137.2	465.8	2751.3	2227.9	
i=3	*-om*	11793.5	8048.5	116.2	17.8	17.2	
i=4	*-im*	4097.7	1800.2	321.4	18.7	13.2	
i=5	*-i*	2262.1	2232.9	2512.5	232	36	
i=6	*-ih*	177.2	75.1	215.5	3240.3	1271.6	
i=7	*-e*	1406	1733.2	7770.8	1351.9	1815.2	
i=8	*-a*	1062.2	513.4	3955.9	227.7	63.1	
i=9	*-oj*	9235.8	3491	24.9	13.1	12.3	
i=10	*-u*	2105.2	1821.4	3649.1	25.2	14.2	
i=11	*-o*	5425.3	2993.2	4915.1	1551.6	297.4	*f(e)*=103932.7
		$q(ij)=f(e_i r_j)/f(e)$					
i=1	*- ø*	0.01199	0.01375	0.01195	0.00211	0.00066	
i=2	*-og*	0.00165	0.00132	0.00448	0.02647	0.02144	
i=3	*-om*	0.11347	0.07744	0.00112	0.00017	0.00017	
i=4	*-im*	0.03943	0.01732	0.00309	0.00018	0.00013	
i=5	*-i*	0.02177	0.02148	0.02417	0.00223	0.00035	
i=6	*-ih*	0.0017	0.00072	0.00207	0.03118	0.01223	
i=7	*-e*	0.01353	0.01668	0.07477	0.01301	0.01747	
i=8	*-a*	0.01022	0.00494	0.03806	0.00219	0.00061	
i=9	*-oj*	0.08886	0.03359	0.00024	0.00013	0.00012	
i=10	*-u*	0.02026	0.01752	0.03511	0.00024	0.00014	
i=11	*-o*	0.0522	0.0288	0.04729	0.01493	0.00286	
		$p(ij)\log(p(ij)/q(ij))$					
i=1	*poznat- ø*	0.07975	0.00244	-0.00072	0.01162	0.01972	
i=2	*poznat-og*	0.00475	0.00613	0.02391	-0.00301	-0.00305	
i=3	*poznat-om*	-0.01618	-0.01196	-0.00017	-0.00003	0.00439	
i=4	*poznat-im*	-0.00324	0.00168	-0.00049	-0.00003	0.00106	
i=5	*poznat-i*	0.01773	-0.00194	-0.00193	0.00326	0.00501	
i=6	*poznat-ih*	0.00997	0.00227	-0.0003	0.02017	0.01797	
i=7	*poznat-e*	0.00814	0.0001	-0.01114	-0.0009	0.00312	
i=8	*poznat-a*	0.03277	0.00236	-0.00606	0.00232	0.01055	
i=9	*poznat-oj*	-0.01312	-0.00534	-0.00004	-0.00002	0.0011	
i=10	*poznat-u*	-0.00321	-0.00093	-0.00556	0.00021	0.00183	
i=11	*poznat-o*	0.01416	-0.00221	-0.00746	-0.00185	0.01105	*D(p||q)*=0.219

1.3.2. Discrimination based predictors

Baayen et al. (2011) showed that a simple model based on the principles of discrimination learning can serve as an explanatory framework for understanding inflectional paradigms. In their analysis, relative entropy was proportional to the summed activation of all the cues (bigraphs) that constitute a given inflected form. Recently, this model was developed further, and several discrimination-based quantifications have been proposed and attested as predictors of recognition latencies (Milin et al., 2017)

To shed more light upon adjective word form processing we ran an independent statistical modelling exercise with the discrimination-based predictors initially proposed by Milin, et al. (2017). The aim of this analysis is to provide a perspective on adjective lexical decision that complements the one based on lexical-distributional predictors (such as frequency and word length) and information-theoretic predictors (i.e., paradigmatic and syntagmatic relative entropy). In particular, this analysis is expected to provide important insights into how a discrimination learning framework can help us understand our decision behaviour in this task.

Working details of the particular type of two-layer network that we applied are explained in Baayen et al. (2011), Baayen, Milin, & Ramscar (2016), and Milin et al. (2017). The last two studies made use of two independently trained discrimination networks: a grapheme-to-lexome (G2L) and a lexome-to-lexome (L2L) network, where, as noted earlier, the term 'lexome' was defined as a pointer to locations in a distributed semantic space. As such, lexomes are a useful representational construct for gauging the role of discriminative learning on the association strengths between sub-lexical orthographic features (i.e., graphemes) and contextually distributed 'meanings'.

Milin et al. (2017) used the Rescorla-Wagner equations to build a G2L network using letter triplets as input cues and lexomes as outcomes, and an L2L network using lexomes as both cues and outcomes. Following this approach we also used letter trigraphs as input cues and word forms as outcomes in building a G2L network, For example, the Serbian adjective word form *poznatog* (Eng. familiar, famous) would form one independent outcome with its own trigraph cues *#po*, *poz*, *ozn*, *zna*, *nat*, *ato*, *tog*, *og#* (# denotes the space character). However, following Milin et al. (2017), we also used all trigraphs that were present in the given utterance where the outcome word occurred. For example, in a sentence "Video je poznatog glumca" (Eng. [He] saw a famous actor), the trigraph cues from preceding and following words *(#je*, *je#*, *e#p*, *#po*, *poz*, *ozn*, *zna*, *nat*, *ato*, *tog*, *og#*, *g#g*, *#gl*, *glu*, *lum*, *umc*, *mca*, *ca#*) would all compete for the discrimination of the outcome adjective word *poznatog*.

In addition to the G2L network, we also built an L2L network, again following the procedure described by Milin et al. (2017). This network represents each lexome as the vector in the space of other lexomes, and implements the previously described definition of the lexome as the "pointer to a location in a high-dimensional co-occurrence based semantic space" (Milin et al., 2017; p. 10).

A Rescorla-Wagner network was trained through iterative exposure to utterances from a 65.5 million word corpus of Ebart Media Database (http://www.arhiv.rs). Three-word sequences were taken as a discrete event to learn association weights between trigraph cues and lexome outcomes. The discrimination learning of the network is summarized in a $k \times n$ matrix of discrimination weights, with k cues and n outcomes (the G2L matrix). The same was performed for learning association weights in a matrix of n input and n output lexomes (the L2L matrix). However, despite the fact that the NDL is dealing with the same sample of n lexomes that form the $n \times n$ L2L matrix, their role in learning is two-fold: once as predictor (a cue in n rows) and once as predictee (an outcome in n columns).

Weights served as a basis to derive discrimination-based predictors of reaction time latencies from our experiment. Following Milin et al. (2017) we targeted three discrimination-based indicators derived from the G2L network: (i) an adjective lexome's activation, representing its bottom-up support (G2L-Activation), (ii) the input cue diversity as a measure of uncertainty and/or competition among outcomes (G2L-Diversity), and (iii) a lexome's long-term availability irrespective of any (perceptual) input (G2L-FormPrior). In addition to these three measures we also derived a measure of the availability of the whole paradigm by taking into account the availability of all inflected forms of a given lexome (G2L-ParadigmPrior). Finally, based on the L2L network, we derived a measure of semantic typicality of the whole adjectival paradigm (Paradigm Typicality).

1.3.2.1.G2L-Activations

As stated earlier, cells of the G2L matrix contain discrimination (associative) weights from *k* trigraph cues to *n* lexome outcomes. For the *j*-th lexome, its activation a_j is defined as

$$a_j = \sum_{i \in C} w_{ij}$$

summing over all active cues (elements of *C*). Typically, the set of active input cues are those that are visually or auditorily presented in the perceptual input (e.g., all trigraphs that are present in an adjective word form). This measure, initially applied in Baayen et al. (2011), has been shown to correlate with empirically observed processing latencies. If treated as the simulation of reaction time, it can account for the various empirical phenomena noted above, including the effect of N-gram frequency (Baayen, Hendrix, & Ramscar, 2013), paradigmatic relative entropy (Filipović Đurđević & Gatarić, in press), morphological family size (Baayen et al., 2011) etc. Non-zero values of G2L-Activations indicate the presence of bottom-up evidence for a given lexome.

1.3.2.2.G2L-Diversity

G2L-Diversity is obtained as the 1-norm given the same active input cues' weights to all outcomes (i.e., 1-norm over rows of the Rescorla-Wagner matrix corresponding to active letter trigraphs):

$$D_C = \sum_{i \in C, j} |w_{ij}|$$

In this equation, we sum over all active cues (elements of *C*) and all *j* columns (outcomes), to get the absolute length of the activation vector. As Milin et al. (2017) explained, "[the] 1-norm highlights mathematically the extent to which there are lexomes that are relevant given the input" (p. 13). At the same time, lexomes that are irrelevant would have approximately zero-weights and, consequently, would not contribute to the 1-norm. Thus, this measure indicates the level of uncertainty at the output level and stands as a plausible proxy for a measure of competition among lexomes.

1.3.2.3.G2L-FormPrior

G2L-FormPrior was calculated as 1-norm of a column vector representing the target lexome (e.g., an inflected adjectival form). This vector places the lexome in the orthographic space of all possible orthographic cues, regardless of the cues that are available in the input. G2L-FormPrior represents the absolute length of this vector and indicates the prior availability of a given lexome, based on prior experiences:

$$P_O = \sum_i |w_{iO}|$$

summing over all rows of the matrix for the lexome outcome – *O*. Studies so far show that this measure is highly correlated with frequency counts.

1.3.2.4.G2L-ParadigmPrior

For the G2L-ParadigmPrior 1-norm was calculated over all columns for inflected forms of the paradigm to which a given lexome belongs. In other words, it encapsulates the full inflectional paradigm of the lexome, and indicates the availability of the inflectional forms regardless of the input:

$$P_{O-O} = \sum_{i,j \in P} |w_{ij}|$$

where we account for all inflected forms of a given adjective paradigm *P*. In terms of traditional measures, this measure could be related to lemma frequency (i.e. cumulative frequency of inflected form frequencies).

1.3.2.5.Paradigm Typicality

Paradigm Typicality is calculated as the cosine similarity of the average column vectors pertaining to all inflected forms of the given adjective and the average column vectors pertaining to the whole L2L network. First, we form a paradigm column vector (as row-wise averaged weights): $\mathbf{P} = \frac{1}{n}\sum_{j \in P} w_{.j}$, where n is the number of columns in the paradigm *P*. Next, we calculate the average column vector of the L2L-matrix: $\mathbf{L2L} = \frac{1}{N}\sum_j w_{.j}$, for *N* columns in the matrix. Finally, we use the two previously-obtained vectors to estimate the typicality as their cosine similarity:

$$\cos(\theta) = \frac{\mathbf{P} \cdot \mathbf{L2L}}{||\mathbf{P}||\ ||\mathbf{L2L}||}$$

The Paradigm Typicality measure serves as an index of semantic typicality, or the semantic “non-remarkableness” of the given adjective as predicted from many co-occurring words in context.

2. Method

2.1.Participants

A total of 155 students from the Department of Psychology at the Faculty of Philosophy, University of Novi Sad and the Department of Psychology at the Faculty of Philosophy, University of Belgrade participated in the study as part of their course requirements. All were native speakers of Serbian and had normal or corrected-to-normal vision. They were randomly assigned to one of the six experimental conditions (26; 25; 25; 28; 26; 25). The participants signed an informed consent form prior to the study.

2.2.Materials and design

We presented 106 adjectives and 106 pseudo-words of Serbian. Adjectives were retrieved from the Frequency Dictionary of the Contemporary Serbian Language (Kostić, 1999), so as to have as high as possible non-zero form frequencies within a paradigm (in the final set they had at least 10 of possible 11 inflected forms with non-zero frequencies and at least 25 of possible 36 cases with non-zero frequencies). The frequencies of the nominative case were joined with those of the (nondistinct) vocative case, as suggested by Kostić (1965). For the purposes of relative entropy calculations, zero values were replaced with 0.1. We excluded adjective-verb homographs, to ensure that the final list consisted solely of adjectives (as listed in Appendix A).

All stimuli were presented in six separate experimental sessions, based on the manipulation of inflected form and the presentation condition (blocked, mixed). In half of the sessions, inflected forms were presented in separate blocks (one inflected form per participant), whereas in the other half, inflected forms were mixed (all participants saw all of the tested inflected forms). In the first session of the blocked design we presented the inflected form marked by an exponent *–og* (e.g. *novog* [new]), denoting the genitive singular of the masculine and neuter gender and the accusative singular of the masculine gender. In the second blocked session the exponent *–om* (e.g. *novom*) was presented, marking the dative and locative singular of the masculine and neuter gender, as well as the instrumental singular of the feminine gender. In the third blocked session we presented the exponent *–oj* (e.g. *novoj*), marking the dative singular of the feminine gender. In the remaining three sessions we presented all of the forms in a single list. The three sessions of the mixed presentation condition were introduced for the purposes of counterbalancing the inflected forms across the stimuli by using a Latin square design. In each session the exponents attached to pseudo-words mirrored those attached to adjectives. The blocked versus mixed presentation condition was introduced to control for potential list effects.

In addition to the variable of inflected form (ambiguous gender – masculine/neuter [-*og*], ambiguous gender – masculine/feminine/neuter [-*om*], feminine gender [-*oj*]), our main independent variables were drawn from a group of information-theoretic measures (paradigmatic relative entropy and syntagmatic relative entropy) and from a group of discrimination learning based measures (G2L-Activations, G2L-Diversity, G2L-FormPrior, G2L-ParadigmPrior, and Paradigm Typicality), as described in the introduction. Adjectival frequencies were retrieved from Frequency Dictionary of Contemporary Serbian Language (Kostić, 1999), and preposition-adjective phrasal frequencies were calculated based on co-occurrences observed in the Ebart Media Database (with more than 65.5 million words of Serbian language; http://www.arhiv.rs). Phrasal frequencies were calculated for all selected adjectives, both with and without the large set of 73 typical Serbian prepositions immediately preceding the inflected adjectival form. Additionally, several variables were included in the regression model as covariate controls (all listed in Appendix B): presentation condition (blocked, mixed), word length in letters, (log) form frequency, (log) lemma frequency, and order of trial presentation. The dependent variable was reaction time.

2.3.Procedure

Participants were presented with a visual lexical decision task. In each trial, a blank screen appeared for 500ms, followed by a fixation point that appeared for 1000ms, followed by a stimulus that remained on the screen until the response or time-out of 1500ms. Participants responded by pressing the left mouse-button with their right index finger for words or the right mouse-button with their right middle finger for pseudo-words. In the case of error or time-out, participants received feedback. Prior to the experiment, 10 practice trials were presented that were not analysed. The order of presentation was randomized individually for each participant. The presentation of stimuli was controlled by OpenSesame experimental software (Mathôt, Schreij, & Theeuwes, 2012).

3. Results

Prior to analysis, we excluded participants with an error rate exceeding 25% and stimuli that induced an error rate exceeding 25%. In total, we excluded nine participants and eight word items. We standardized all numeric predictors (by centring to zero and dividing by the standard deviation – z-scores, as suggested by Gelman & Hill, 2007). Reaction time was transformed by applying reciprocal transformation (-1000/RT), following Baayen and Milin (2010).

The data were analysed in **R** (http://www.r-project.org/), using **mgcv** (Wood, 2006; 2011), **itsadug** (van Rij, Wieling, Baayen, & van Rijn, 2016), and **gbm** (Ridgeway et al., 2017) packages. We fitted a Generalized Additive Mixed Model to reaction time latencies using Gaussian distribution as the appropriate functional form. In addition to testing the fixed effects, we controlled for two random effects. The random effect of words was taken into account by including by-item adjustments for word forms. The random effect of participants was dealt with by including by-participant factorial smooths over experimental trials, which accounted for fluctuations that could originate in fatigue, attentional slips, or adaptation to the task. Importantly, by bringing by-participant factor smooths into the model we took care of the inter-trial dependencies in the response latency time-series that occurred in the particular randomized trial sequence to which a participant was exposed. After the initial fit, we excluded standardized residuals that fell outside the -/+2.5 range, and then refitted the model. This model criticism proved that the final model is robust, as there were no substantial differences between the original and the models with influential residual values removed.

We proceed now to a discussion of the re-fitted model.

3.1. Information theory based predictors

The final model in which we fitted lexical distributional and information-theoretical predictors is presented in Table 3 and in Figure 1. In addition to a significant non-linear inhibitory effect of word length, we observed significant a three-way interaction of lemma frequency, paradigmatic relative entropy, and syntagmatic relative entropy which we will discuss in detail below. As Figure 1 reveals, we see a fine interplay among the three predictors. First and foremost, both measures of relative entropy exhibit the predicted inhibitory effect on processing time. However, the strength of this effect is dependent on lemma frequency. The syntagmatic relative entropy effect is at its strongest when paradigmatic relative entropy and lemma frequency are at their lowest (Figure 1, leftmost panel). This effect gradually attenuates as lemma frequency increases (Figure 1, panels from left to right, showing upward trends on the left-hand side). At the same time, as lemma frequency increases, paradigmatic relative entropy becomes more and more prominent (Figure 1, panels from left to right, horizontal trends for low and mid values of the Y-axis that is representing syntagmatic relative entropy). The paradigmatic relative entropy is most prominent for the words with the highest lemma frequency and where syntagmatic entropy distribution is the densest (as most adjectives exhibit close-to-average values of syntagmatic relative entropy, which is presented on Y-axis).

Finally, the panels on Figure 1 show large triangular-shaped white patches in the upper-right corners. This means that we were not able to observe *any adjective* that is *simultaneously highly divergent syntagmatically and paradigmatically*. Without drawing excessively bold and far-reaching conclusions, we find this a rather intriguing manifestation of adjectival complexity which certainly deserves further scrutiny, and which we will return to in subsequent sections of the manuscript.

Table 3. Coefficients from the generalized additive mixed model of lexical-distributional predictors fitted to transformed response latencies.

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept (Blocked presentation)	-1.524	.026	-57.771	<.0001
Order of trial presentation	-.004	.006	-.703	0.482
Presentation Condition (Mixed)	-.144	.036	-4.014	<.0001

Smooth terms:

	edf	Ref.df	F	p-value
Smooth for Word Length (in characters)	1.965	1.972	35.198	<.0001
Tensor product smooth for (log) Lemma Frequency, Paradigmatic Relative Entropy, and Syntagmatic Relative Entropy	15.423	15.914	5.316	<.0001
By-Participant factor smooths for Order of Trial Presentation	540.550	1339.000	7.906	<.0001
By-Item random intercept	227.722	299.000	3.530	<.0001

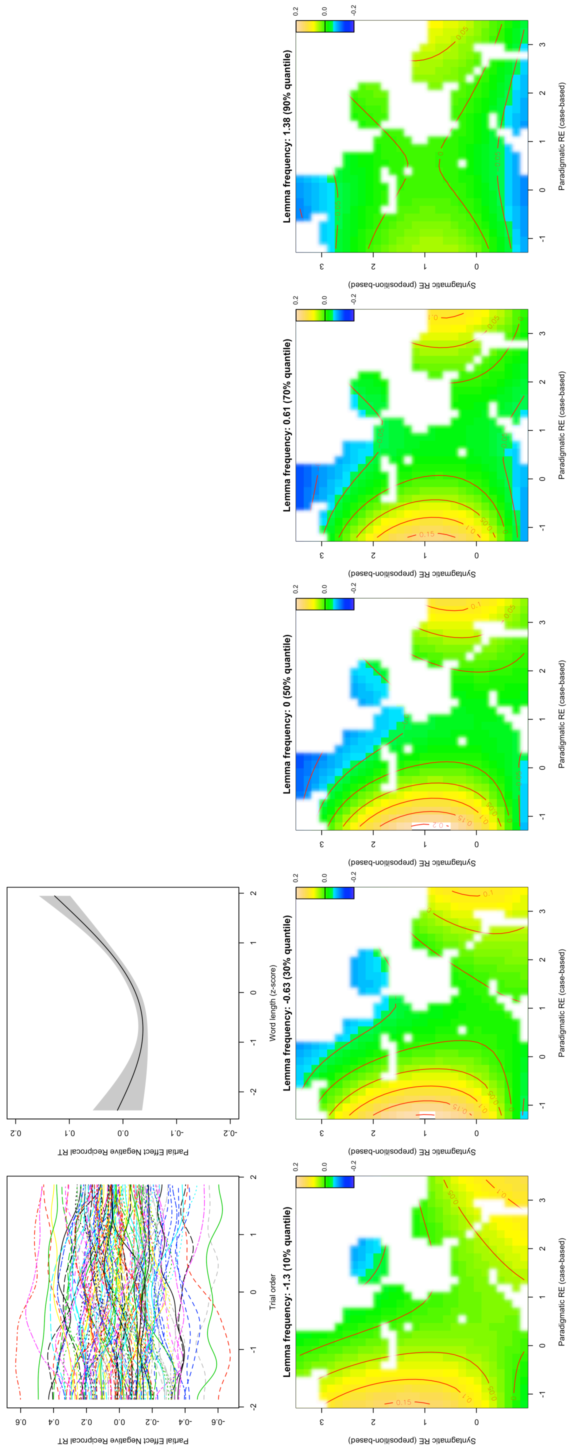

Figure 1. Generalized additive model fitted to transformed response latencies; upper row: partial effects of by-participant smooths for the order of trial presentation and smooth for word length (in characters); bottom row: tensor product smooth for (log) lemma frequency, paradigmatic relative entropy, and syntagmatic relative entropy; colours indicate variation in transformed response latencies.

3.2.Naïve Discrimination Learning (NDL) based predictors

Discrimination weights are distributed symmetrically but are exceptionally spiky (which can be described as a Generalized Hyperbolic distribution, or as its special case – the Normal-Inverse Gaussian distribution, as discussed in Milin et al., 2017). Hence, all discrimination based predictors (G2L-Activation, G2L-Diversity, G2L-FormPrior, G2L-ParadigmPrior, and Paradigm Typicality) were rank-transformed to standardized normally-distributed values (c.f., Johnson, 1949; Chou, Polansky, & Mason, 1998) to facilitate statistical modelling.

Prior to fitting a statistical model to reaction time, we performed a descriptive analysis and ran collinearity diagnostics on our predictor set. The matrix of correlations between two sets of predictors – lexical-distributional/information-theoretic vs. discrimination-based – revealed some significant relations, mainly weak-to-moderate (see Table 4, representing correlations > .19, which were all at significance level of $p < 0.001$). Most of those significant correlations are concentrated between lexical-distributional predictors (lemma frequency and word length) and discrimination-based predictors, with one distinctively higher coefficient ($r = .897$) between word length and cue diversity. This suggests that longer adjectives simply have more trigraphs which, consequently, can induce stronger competition (as indicated by G2L-Diversity), even if by pure chance. Here we are not referring to the competition with orthographic neighbours as traditionally defined via number of words that can be obtained by replacing a single grapheme (the count of which is indeed smaller for long words). Instead, we refer to the competition of outcomes that are "fed" by individual orthographic cues (trigraphs) that constitute a given input word (and the number of which is simply larger for long words). Longer words have more trigraphs and this increases the chance that some will occur in a number of other words, which is then reflected in respective Diversity.

We also observed a positive correlation between G2L-ParadigmPrior and (log) lemma frequency (.244). Baayen, Milin, & Ramscar (2016) and Milin et al. (2017) argued that, in fact, G2L-Prior reflects the effect of frequency in discrimination learning. Importantly, in discrimination learning the effect of frequency is captured not only by positive evidence, but also by the instances when outcomes are absent (as mentioned in 1.2, the presence of the cue(s) which is not accompanied by the presence of the target(s) leads to a decrease in connection weights for the target(s) in question). Although originally related to G2L-FormPrior, this interpretation can be applied to G2L-ParadigmPrior, as well. Syntagmatic relative entropy was also negatively correlated with G2L-Activation, G2L-ParadigmPrior, G2L-FormPrior, and ParadigmTypicality. This suggested that inflected adjectives with an unusual pattern of prepositioning tended to produce less activation, were less entrenched in the system, and had a less typical activation profile, i.e., were contextually unusual.

Table 4. Significant correlations ($r > 0.19$; $p < 0.001$) between lexical-distributional/information-theoretic and discrimination-based predictors.

	G2L-Activation	G2L-Diversity	G2L-Paradigm Prior	G2L-Form Prior	Paradigm Typicality
Word length (in characters)		.897			
(log) Lemma Frequency		-.191	.244	.194	
Paradigmatic Relative Entropy				.194	
Syntagmatic Relative Entropy	-.271		-.298	-.294	-.286

Correlations between the four discrimination-based predictors also showed some strong relationships, particularly between G2L-FormPrior and G2L-Activation (r = .917), and between G2L-FormPrior and G2L-ParadigmPrior (r = .779). G2L-FormPrior was also one of the main suspected sources of collinearity issues in this set of predictors, with Variance Inflation Factor VIF = 11.146 (by comparison, the collinearity condition number κ shrunk more than three times after removing this measure, from 10.249 to 2.806). For these reasons we discarded G2L-FormPrior from further analyses.

Next, a variable importance analysis, using a Generalize Boosting Regression Modelling as implemented in **gbm** package in **R** (Ridgeway et al., 2017), showed higher importance of the discrimination based predictors as compared with the importance of lexical-distributional/information-theoretic predictors, for decision time latencies in our experiment (Figure 2). G2L-Diversity appeared to be the single most important predictor, followed by the group of three candidate predictors: G2L-ParadigmPrior, word length, and Paradigm Typicality. Paradigm-based relative entropy, lemma frequency, G2L-Activation, and Syntagmatic Relative Entropy follow as the group of predictors with the lowest, mutually similar importance.

This concludes our preliminary analyses and we now proceed to model reaction time latencies that engage discrimination-based predictors.

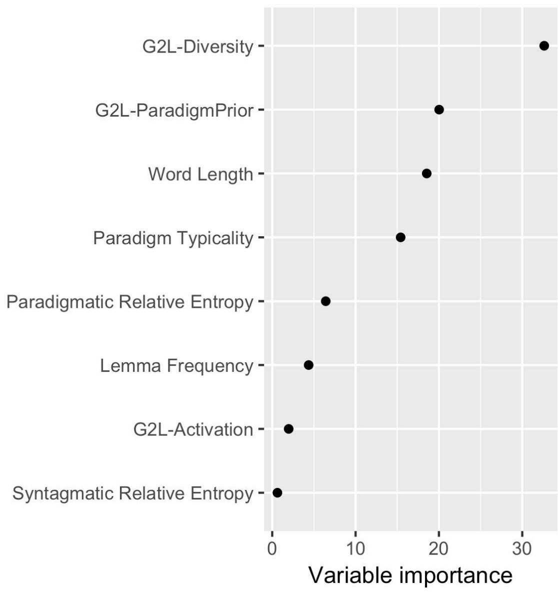

Figure 2. Relative importance of predictors, as revealed by Generalized Boosting Modelling.

We fitted a Generalized Additive Mixed Model to processing latencies. The model revealed a significant interaction of G2L-Diversity and Paradigm Typicality. However, this interaction was additionally modulated by the G2L-ParadigmPrior. The discrimination-based model gave a better statistical fit as compared to the model based on lexical-distributional and information-theoretic measures. The comparison between the lexical-distributional/information-theoretic model and the discrimination-based model showed a $fREML$ difference of 5.546 ($p < 0.0001$).

Table 5. Coefficients from the generalized additive model of discrimination-based predictors fitted to transformed response latencies.

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept (Blocked presentation)	-1.523	.026	-57.796	<.0001
Order of trial presentation	-.004	.006	-.730	.466
Presentation Condition (Mixed)	-.144	.036	-4.022	<.0001

Smooth terms:

	edf	Ref.df	F	p-value
Smooth for Word Length (in characters)	1.964	1.971	21.701	<.0001
Tensor product smooth for G2L-ParadigmPrior, ParadigmTypicality, and G2L-Diversity	10.665	10.880	7.488	<.0001
By-Participant factor smooth for Order of Trial Presentation	539.057	1339.000	7.887	<.0001
By-Item random intercept	232.588	299.000	3.673	<.0001

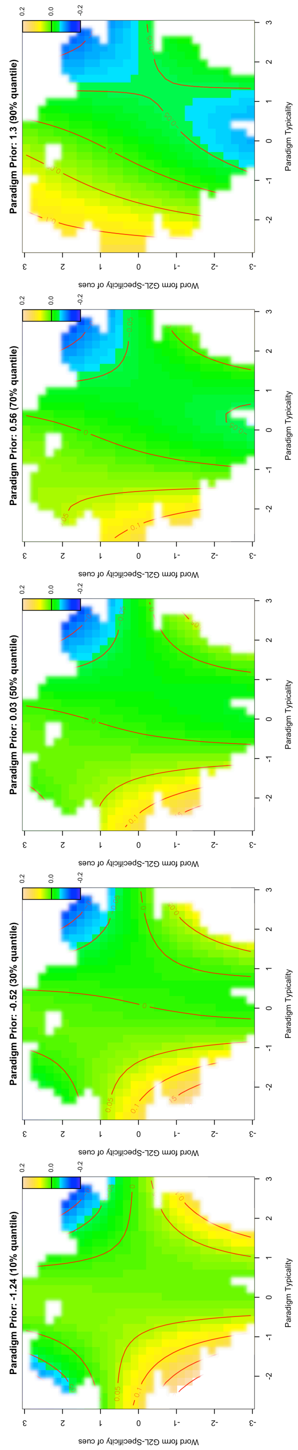

Figure 3. Generalized additive model fitted to transformed response latencies with predictors derived from Naïve Discrimination Learning framework – Paradigm Prior, Paradigm Typicality, and Specificity of cues, the latter being the additive inverse of the Diversity, plotted as such for the purposes of comparison with Figure 1; colours indicate variation in transformed response latencies.

The critical three-way interaction of Paradigm Typicality with G2L-Diversity with G2L-ParadigmPrior is presented on Figure 3. (Here, we did not plot the effect of word length because it shows a remarkable similarity with the effect presented on Figure 1.) For the purposes of comparison with the model reported in 3.1., we present G2L-Diversity as an additive inverse (i.e., by multiplying its values with -1) and label it, for consistency, as G2L-Specificity.

We observe the effect of both Specificity of cues and ParadigmTypicality across different values of ParadigmPrior (presented as separate panels). The dynamics between the two changes as ParadigmPrior values increase. The effect of G2L-Specificity is facilitatory for adjectives with lower entrenchment of the inflected paradigm forms (i.e., G2L-ParadigmPrior, compare changes across panels 1-4 on Figure 3). This effect attenuates and Paradigm Typicality takes over as a facilitation effect for the adjectives with the strongest Paradigm Prior (compare panels 4-5 on Figure 3). In addition, when G2L-ParadigmPrior is at its weakest (panels 1 and 2 on Figure 3), we find facilitatory effect of Specificity for cues for words that have below average (approximately <-1) values of Paradigm Typicality, and words that have above average (approximately >1) values. Taken together, we may argue that the Specificity of cues makes the decision easier (much as the Diversity of cues makes the decision harder, as demonstrated by Milin et al. 2017). However, the absence of competition incurred by Specificity of cues will be attenuated as the support from the long-term experience increases, which is when and where we see the greatest processing support attributable to ParadigmTypicality (i.e., how typical a given adjectival paradigm is).

This pattern of effects resembles, to a certain extent, the three-way interaction of lemma frequency and two relative entropies (five lower row panels on Figure 1). In order to formalize the comparison of the two models, we looked at the squared correlation coefficient of the predicted values obtained by the two models. After partializing out random effects, the predictions derived from the two models share an astounding 85.83% of variance (i.e., $r^2 = 0.858$), thus confirming the parallel that we discussed based on the observed effects. The two models, however, offer complementary perspectives on the processing complexity of adjectival paradigms, providing a more comprehensive picture of the processing of adjectives in a morphologically rich language.

4. General discussion

Our results showed a strikingly complex pattern of effects which predicted the time spent in deciding whether a visually presented string was a Serbian word form or not. The first statistical model, the one that tested a set of lexical-distributional and information-theoretic based predictors, shows that processing latency is co-determined by syntagmatic and paradigmatic properties of the given inflected adjectival form. This synergy, which is captured by the interaction between two relative entropies – two indicators of a word's usage complexity, is further modulated by lemma frequency. Adjectives with low-frequency and atypical prepositional realization require additional decision-making time; as lemma frequency becomes higher, the syntagmatic complexity effect attenuates and is overtaken by its paradigmatic counterpart. The two relative entropy effects show inhibitory tendencies which indicate that the more syntagmatically or paradigmatically atypical an adjective is, the more demanding the processing and, consequently, the harder the decision task becomes.

When it comes to the effect of paradigmatic relative entropy, our finding generalizes the effects that have been found in other word classes that exhibit nominal inflection. The present results thus support the conclusions suggested by Milin et al.'s (2009) study of the processing of Serbian nouns, and reinforced by similar findings observed with visually presented Serbian verbs (Filipović Đurđević & Gatarić, in press), English nouns (Baayen, et al., 2011; Hendrix & Baayen, 2014; Kuperman et al., 2010; Linzen, et al., 2013; Milin, Kuperman et al., 2009), and auditorily presented Romanian verbs (Nenadić, et al. 2016). Taken together, these findings implicate a single principle that favours minimal divergence between the distribution of a paradigm and the distribution of its class. This principle, which appears to apply to different PoS categories and even across languages, encodes sensitivity to a word's 'conformism' (i.e., prototypicality) in its use in a wider linguistic system.

As far as the effect of the syntagmatic relative entropy is concerned, we have argued that case/number ambiguity is resolved contextually, within the prepositional phrase which contains an adjective. The complexity of this prepositional phrase was captured by the use of syntagmatic relative entropy. The observed importance of syntactic complexity suggested that the cues for various grammatical features seem to be distributed across the successive units of utterance. If we generalize this to the wider linguistic context, we would say that, as the communication unfolds, the preposition clears the way for the adjective which, further, increases the expectations for the following noun.

A similar uncertainty-reducing role of gendered prenominal determiners in German is discussed by Dye, Milin, Futrell, and Ramscar (2017). The same authors demonstrated that gender markings in German serve the same role as prenominal adjectives in English (in Dye, Milin, Futrell, & Ramscar, 2018). However, whereas their findings referred to the identity of the noun (i.e. its semantics), we are focusing on the morpho-syntactic features of a word. In a complex inflectional system, such as Serbian, the syntactic ambiguity of the inflected form is more easily resolved if some of its grammatical features are previously disambiguated by the preceding words. For example, the disambiguated gender can reduce uncertainty about the inflectional paradigm/class, which then reduces the uncertainty about the grammatical case/number. Similarly, the disambiguated grammatical case/number can reduce uncertainty about the grammatical gender, and so on. However, uncertainty is not fully resolved by any of the prenominal words. Instead, during the course of communication, information flow is enabled by the very balancing of the uncertainty (c.f., Shannon, 1948; Ramscar & Baayen, 2013). At the point when a noun is encountered, the uncertainty about the gender (as well as grammatical case/number) is fully resolved and the communication can proceed navigating through further uncertainties. Any word – an adjective in this particular case – plays a role in reducing/has as its raison d'être the reduction of uncertainty in communication, and our present results convincingly show how complex this process must be; i.e., *how converging communicative forces remain in balance to achieve efficiency in message transmission*.[5] Importantly, it should be noted that our task did not involve successive presentation of phrasal constituents, nor the phrases themselves. We presented isolated inflected adjective forms, and yet observed the effects of the usage of the particular form across prepositional phrases relative to the average use of other adjectives in that inflected form across the same prepositional phrases. Similar remarks apply to the effect of paradigmatic entropy.

In addition to demonstrating the effects of paradigmatic and syntagmatic entropy, we have shown how the two act simultaneously. Moreover, we showed how their synergy is

[5] Our accidental but striking finding – that we were not able to attest a single adjective that was highly atypical both syntagmatically and paradigmatically at the same time – could be seen as the exact reaffirmation of this point: language is complex adaptive system that may allow uncertainty peaks in one or another dimension but not in all, simultaneously. However, before building further on this finding, more research is needed to reject an alternative explanation that this double atypicality might also be a simple consequence of the mutual independence of the paradigmatic and syntagmatic atypicality (as the product of two low probabilities gives even lower probability value).

coupled with lemma frequency, i.e. how the importance of each is amplified by lemma frequency. This is exactly what the discrimination-based model emphasized by revealing different but equally intriguing reciprocation of various effects on lexical decision making: the adjectives that are not learned well will be ‘punished’ when input cues are co-activating multiple potential lexomic competitors. However, this drawback could still be compensated for if those cues, including the ones from the preceding words (such as prepositions), provided support to an outcome which is in the right range of typicality of meaning (i.e. it neither denotes a highly remarkable nor a highly neutral meaning). For well-entrenched (well-learned) adjectives, the cue competition (i.e., diversity) will be beneficial during decision making and no further support is needed.

From the Naïve Discrimination Learning framework, paradigmatic effects are a consequence of the process of discrimination. Even though our main finding could be interpreted as simultaneous activation of all members of the given paradigm or class, an NDL perspective offers a simpler explanation, and one that does not require that all encountered exemplars be stored in memory (c.f., Baayen et al., 2011). This explanation reflects the fact that the elements of language, as traditionally conceived, are treated in NDL only as links between linguistic input (e.g. bigrams, trigrams, acoustic features) and knowledge of the world (which is distributed via weights of connections among cues and outcomes). From the standpoint of the NDL approach, the advantage of words with multiple morphological relations (in this case syntagmatic as well as paradigmatic) is expressed by a richer set of connections between forms and contextually distributed ‘meanings’. Words with such abundant morpho-semantic relations have higher chances to be learned discriminatively – a consequence of being experienced across consistent yet diversified inflectional variants. Hence, the system is tuned to learn systematic discriminative mappings between forms and slight changes in their syntactic realizations (functions and meanings) of essentially the same lexome (Baayen et al., 2011; Baayen, Hendrix, & Ramscar, 2013; Hendrix et al., 2016). A further implication is that in languages that do not exhibit much morphological complexity, the system would achieve similar *discriminative potential* using the other linguistic means at its disposal (cf., Dye, Milin et al., 2017a; Dye, Milin et al., 2017b).

With respect to the relation between the set of measures derived from information theory and those derived from NDL, it could be argued that both capture the same probabilistic traits of language but engage with a (seemingly) different set of indicators. And in this study, the learning-based measures come out as the more convincing predictors, both in terms of the variable importance from Gradient Boosting Machines, and in terms of overall model fit from Generalized Additive Mixed Models. Although it is beyond the scope of this paper to go into the intricacies of the structural relationships between these two sets of measures, it is important to point out that both sets of measures may be complementary rather than mutually exclusive. The information-theoretic measures indicate a processing challenge in a specific part of the language system, while the learning-based measures are informative about how that challenge might be addressed, in particular, by biologically (i.e., cognitively) sound mechanisms.

In sum, this study combines information-theoretic and discrimination-based approaches to achieve in-depth understanding of the processing of inflected words. Whereas the former highlights important points about the complexity of the process itself, the latter portrays how all of the parts of the “language game” (in Wittgenstein’s sense, 1953, p. 15) could fall in place at the time of lexical decision making. Despite the fact that in the present task the ‘message’ was decontextualized, a carrier-system such as language and a carrier-mechanism such as learning are revealing how they manage to bring this rich dynamics to life: by discriminating for the sake of making minimal errors while “reproducing at one point either exactly or approximately a message selected at another point” (Shannon, 1949, p. 31).

5. Conclusions

To summarize, this research brought several insights. Firstly, we demonstrated that the processing of isolated inflected adjectival forms was influenced by the divergence of the frequency distribution of the inflectional paradigm from the frequency distribution of the inflectional class (as operationalized via relative entropy). The more atypical the paradigm distribution is, the longer it takes to make the lexical decision for the presented inflected form. This is in accordance with the finding previously observed with nouns and verbs, thus revealing that the same principle applies across different parts of speech, modulo certain category dependencies (such as sensitivity to grammatical gender and case/number combinations etc.). We also demonstrated the immense complexity of information contained in language and how various parts of it could arise as a consequence of the process of discrimination in the service of efficient prediction of our (social) environment.

6. Funding

This work was supported by Ministry of Education, Science and Technological Development of the Republic of Serbia (grant number: 179033 and 179006).

7. Author note

All correspondence should be addressed to Dušica Filipović Đurđević, Department of Psychology, Faculty of Philosophy, Čika Ljubina 18-20, 11 000 Beograd, Serbia, e-mail: dmfilipo@f.bg.ac.rs; dfdurdevic@ff.uns.ac.rs, tel: +381 (0) 11 2630 542, fax: +381 (0) 11 2639 356

8. Acknowledgements

The authors would like to express their gratitude to Christian Adam from The University of Sheffield whose help in NDL training was substantial. Ksenija Mišić, from the Laboratory for Experimental Psychology, Faculty of Philosophy, University of Belgrade, helped in the process of data collection and we are in great debt to her too. We would also like to thank the two anonymous reviewers whose in-depth comments contributed immensely to the quality of the manuscript, and to the audience and reviewers of the conferences where we presented parts of our research (First International Quantitative Morphology Meeting, July 11-15 2015, Belgrade, Serbia; International Meeting of the Psychonomic Society, May 5-8 2016, Granada, Spain; and Architectures and Mechanisms of Language Processes, September 6-8, 2018, Berlin, Germany). We are also thankful to Dagmar Divjak (The University of Sheffield), whose comments helped the development and articulation of several parts of the manuscript. Last but not least, our deepest gratitude goes to James Blevins (University of Cambridge, UK) why helped us generously in making this manuscript more accessible and consistent throughout.

9. References

Armstrong BC, Frost R, Christiansen MH. 2017. The long road of statistical learning research: Past, present and future. Philosophical Transactions of the Royal Society B: Biological Sciences. 372: 20160047.

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

Aronoff M. 1976. Word Formation in Generative Grammar. Massachussetts: The MIT Press.

Baayen RH, 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 RH, Milin, P. 2010. Analyzing reaction times. International Journal of Psychological Research. 32:12-28.

Baayen RH, Milin P, Filipovic 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.

Baayen RH, Milin P, Ramscar M. 2016. Frequency in lexical processing. Aphasiology. 1-47.

Belsley DA, Kuh, E. & Welsch, R. E. 1980. Regression Diagnostics. Identifying Influential Data and Sources of Collinearity, Wiley Series in Probability and Mathematical Statistics, New York.

Blevins JP. 2003. Stems and paradigms. Language. 79: 737–767.

Blevins JP. 2006. English inflection and derivation. In B. Aarts & A. M. McMahon Eds., Handbook of English linguistics. London, England: Blackwell. p 507–536.

Boston MF, Hale JT, Patil U, Kliegl R, Vasishth, S. 2008. Parsing costs as predictors of reading difficulty: an evaluation using the Potsdam sentence corpus. Journal of Eye Movements Research. 2: 1–12.

Brysbeart M, New B. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods. 414: 977–990. pmid:19897807

Chou YM, Polansky AM, Mason RL. 1998. Transforming non-normal data to normality in statistical process control. Journal of Quality Technology. 302: 133-141.

Cover TM, Thomas JA. 1991. Elements of information theory. New York: John Wiley & Sons.

De Jong NH, Schreuder R, Baayen RH. 2000. The morphological family size effect and morphology. Language and Cognitive Processes. 15: 329-365.

Demberg V, Keller F. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. Cognition. 109: 193–210.

Dye M, Milin P, Futrell R, Ramscar M. 2017. A functional theory of gender paradigms. In F. Kiefer, J.P. Blevins, & H. Bartos Eds. Perspectives on Morphological Organization: Data and Analyses. Brill: Leiden.

Dye M, Milin P, Futrell R, Ramscar M. 2018. Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. Topics in Cognitive Science. doi: 10.1111/tops.12316

Filipović Đurđević D, Gatarić I. In press. Simultaneous effects of inflectional paradigms and classes in processing of Serbian verbs. Psihologija. https://doi.org/10.2298/PSI170811015F

Filipović D, Kostić A. 2003. Kognitivna obrada prideva. Psihologija. 363: 353-378.

Filipović D, Kostić A. 2004. Kognitivni status roda prideva u srpskom jeziku. Psihološka istraživanja. 14: 125-168.

Frank SL. 2013. Uncertainty reduction as a measure of cognitive processing load in sentence comprehension. Topics in Cognitive Science. 5: 475–494.

Frank SL, Otten LJ, Galli G, Vigliocco, G. 2015. The ERP response to the amount of information conveyed by words in sentences. Brain and Language. 140: 1-11.

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

Hale JT. 2001. A probabilistic early parser as a psycholinguistic model. Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics (Vol. 2, pp. 159–166). Pittsburgh, PA: Association for Computational Linguistics.

Hale JT. 2003. The information conveyed by words. Journal of Psycholinguistic Research. 32: 101–123.

Hale JT. 2006. Uncertainty about the rest of the sentence. Cognitive Science. 30: 643–672.

Henderson JM, Choi W, Lowder M, Ferreira, F. 2016. Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in Reading. NeuroImage. 132: 293-300.

Hendrix P, Baayen RH. 2016. Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. Journal of Experimental Psychology: Learning, Memory and Cognition. 43(1): 128-149.

Jakić, M. (2016). Pridevska antonimija u rečniku, kontekstu i kognitivnom sistemu. Unpublished doctoral dissertation. Faculty of Philology, University of Belgrade.

Johnson NL. 1949. Systems of Frequency Curves Generated by Methods of Translation. Biometrika. 36 1/2: 149-176.

Jost E, Christiansen MH. (2016). Statistical learning as a domain-general mechanism of entrenchment. In H.-J. Schmid (Ed.), Entrenchment, memory and automaticity: The psychology of linguistic knowledge and language learning. Boston: APA and Walter de Gruyter. p 227-244

Kostić A. 1991. Informational approach to processing inflected morphology: Standard data reconsidered. Psychological Research. 53(1): 62-70.

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

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. Berlin: Mouton de Gruyter. p 1-44.

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. 1965. Sintaksičke funkcije padežnih oblika u srpskohrvatskom jeziku. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora.

Kostić Đ. 1999. Frekvencijski rečnik savremenog srpskog jezika. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju.

Kuperman V, Bertram R, Baayen RH. 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

Linzen T, Marantz A, Pylkkänen L. 2013. Syntactic context effects in single word recognition: An MEG study. The Mental Lexicon. 82: 117-139. DOI: http://dx.doi.org/10.1075/ml.8.2.01lin

Mathôt S, Schreij D, Theeuwes J. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods. 442: 314-324.

Milin P, Feldman LB, Ramscar M, Hendrix P, Baayen RH 2017. Discrimination in lexical decision. PLoS One. 122. e0171935-e0171935.

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.

Milin P, Kuperman V, Kostić A, Baayen RH. 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. Oxford, England: Oxford University Press. p 214–252.

Moscoso del Prado Martín F, Kostić A, Baayen RH. 2004. Putting the bits together: An information theoretical perspective on morphological processing. Cognition. 94: 1–18.

Nenadić F, Tucker BV, 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.

Ramscar M, Baayen RH. 2013. Production, comprehension and synthesis: A communicative perspective on language. Frontiers in Language Sciences. 4: 233. doi: 10.3389/fpsyg.2013.00233

Ramscar M, Port R. 2016. How spoken languages work in the absence of an inventory of discrete units. Language Sciences. 53: 58-74.

Ramscar M, Yarlett D. 2007. Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. Cognitive Science. 31: 927-960.

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.

Rescorla RA, Wagner AR. 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, Appleton-Century-Crofts. p 64–99.

Ridgeway G, with contributions from others 2017. gbm: Generalized Boosted Regression Models. R package version 2.1.3. https://CRAN.R-project.org/package=gbm

Schmidtke D, Kuperman V, Gagné CL, Spalding TL. 2016. Competition between conceptual relations affects compound recognition: the role of entropy. Psychonomic Bulletin & Review. 23: 556-570.

Shannon CE. 1948. A mathematical theory of communication. Bell System Technical Journal, XXVII: 379-423.

Stanojčić Ž, Popović Lj. 1992. Gramatika srpskog jezika. Beograd: Zavod za udžbenike i nastavna sredstva.

Tiedemann J. 2012. Parallel Data, Tools and Interfaces in OPUS. In: Choukri, K., Declerck, T., Dogan, M.U., Maegaard, B, Mariani, J., et al. editors, Proceedings of the Eight International Conference on Language Resources and Evaluation. Istanbul, Turkey: European Language Resources Association. p 2214–2218.

Widrow B, Hoff ME. 1960. Adaptive switching circuits. 1960 WESCON Convention Record Part IV. p 96–104.

Wittgenstein L. 1953. Philosophical Investigations. Oxford (UK): Blackwell Publishing.

Wood SN. 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Wood SN. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society B 731. p 3-36.

Appendix A

Stimuli presented in the experiment.

Form -OG		Form -OM		Form -OJ	
Word form	Surface frequency (per 2 millions)	Word form	Surface frequency (per 2 millions)	Word form	Surface frequency (per 2 millions)
ATOMSKOG	8	ATOMSKOM	4	ATOMSKOJ	3
BESKRAJNOG	17	BESKRAJNOM	31	BESKRAJNOJ	9
BLISKOG	8	BLISKOM	6	BLISKOJ	3
BOLESNOG	10	BOLESNOM	12	BOLESNOJ	6
BUDNOG	6	BUDNOM	3	BUDNOJ	3
BURNOG	5	BURNOM	15	BURNOJ	4
ČELIČNOG	5	ČELIČNOM	4	ČELIČNOJ	1
ČITAVOG	57	ČITAVOM	25	ČITAVOJ	33
ČUDNOG	13	ČUDNOM	31	ČUDNOJ	10
ČUDESNOG	5	ČUDESNOM	1	ČUDESNOJ	3
ČVRSTOG	16	ČVRSTOM	16	ČVRSTOJ	3
DAVNOG	18	DAVNOM	13	DAVNOJ	3
DESNOG	9	DESNOM	18	DESNOJ	11
DNEVNOG	19	DNEVNOM	16	DNEVNOJ	9
DRUŠTVENOG	64	DRUŠTVENOM	24	DRUŠTVENOJ	3
DUBOKOG	16	DUBOKOM	43	DUBOKOJ	6
EKONOMSKOG	31	EKONOMSKOM	29	EKONOMSKOJ	13
FILMSKOG	5	FILMSKOM	6	FILMSKOJ	5
GLAVNOG	78	GLAVNOM	29	GLAVNOJ	17
GORKOG	22	GORKOM	10	GORKOJ	8
GORSKOG	16	GORSKOM	12	GORSKOJ	4
GRADSKOG	111	GRADSKOM	48	GRADSKOJ	26
GUSTOG	13	GUSTOM	19	GUSTOJ	11
JAKOG	3	JAKOM	43	JAKOJ	3
JUŽNOG	10	JUŽNOM	11	JUŽNOJ	6
KONAČNOG	15	KONAČNOM	1	KONAČNOJ	7
KONKRETNOG	6	KONKRETNOM	4	KONKRETNOJ	1
KRATKOG	20	KRATKOM	29	KRATKOJ	6
KUĆNOG	5	KUĆNOM	16	KUĆNOJ	1
KULTURNOG	26	KULTURNOM	27	KULTURNOJ	14
LAKOG	5	LAKOM	16	LAKOJ	8
LAŽNOG	8	LAŽNOM	6	LAŽNOJ	1
LIČNOG	9	LIČNOM	3	LIČNOJ	3
LJUDSKOG	61	LJUDSKOM	20	LJUDSKOJ	16
MALENOG	3	MALENOM	2	MALENOJ	2
MASOVNOG	8	MASOVNOM	7	MASOVNOJ	1

Form -OG		Form -OM		Form -OJ	
Word form	Surface frequency (per 2 millions)	Word form	Surface frequency (per 2 millions)	Word form	Surface frequency (per 2 millions)
MEKOG	19	MEKOM	22	MEKOJ	21
MIRISNOG	0.1	MIRISNOM	6	MIRISNOJ	3
MOĆNOG	3	MOĆNOM	6	MOĆNOJ	4
MORSKOG	17	MORSKOM	10	MORSKOJ	11
MRAČNOG	11	MRAČNOM	19	MRAČNOJ	13
MUTNOG	26	MUTNOM	23	MUTNOJ	16
MUZIČKOG	14	MUZIČKOM	2	MUZIČKOJ	7
NAUČNOG	10	NAUČNOM	2	NAUČNOJ	5
NEBESKOG	26	NEBESKOM	18	NEBESKOJ	12
NEČUJNOG	3	NEČUJNOM	8	NEČUJNOJ	1
NEMIRNOG	17	NEMIRNOM	14	NEMIRNOJ	15
NEPOZNATOG	21	NEPOZNATOM	19	NEPOZNATOJ	12
NEVIDLJIVOG	18	NEVIDLJIVOM	11	NEVIDLJIVOJ	5
NEVINOG	7	NEVINOM	6	NEVINOJ	3
NEŽNOG	7	NEŽNOM	9	NEŽNOJ	3
NEZNANOG	15	NEZNANOM	11	NEZNANOJ	2
NISKOG	11	NISKOM	9	NISKOJ	8
NOĆNOG	14	NOĆNOM	11	NOĆNOJ	23
OBIČNOG	12	OBIČNOM	9	OBIČNOJ	1
ODLIČNOG	13	ODLIČNOM	29	ODLIČNOJ	3
OGROMNOG	17	OGROMNOM	15	OGROMNOJ	13
OTVORENOG	12	OTVORENOM	16	OTVORENOJ	1
OZBILJNOG	5	OZBILJNOM	5	OZBILJNOJ	1
PLEMENITOG	5	PLEMENITOM	7	PLEMENITOJ	2
POLITIČKOG	38	POLITIČKOM	48	POLITIČKOJ	12
POSEBNOG	8	POSEBNOM	12	POSEBNOJ	2
POVOLJNOG	2	POVOLJNOM	5	POVOLJNOJ	7
POZNATOG	25	POZNATOM	14	POZNATOJ	6
PRAVILNOG	11	PRAVILNOM	7	PRAVILNOJ	1
PRIRODNOG	11	PRIRODNOM	2	PRIRODNOJ	2
PRIVREDNOG	39	PRIVREDNOM	17	PRIVREDNOJ	10
PRLJAVOG	7	PRLJAVOM	10	PRLJAVOJ	8
PROKLETOG	6	PROKLETOM	3	PROKLETOJ	1
RADOSNOG	5	RADOSNOM	7	RADOSNOJ	5
RANJENOG	19	RANJENOM	7	RANJENOJ	0.1
RATNOG	30	RATNOM	5	RATNOJ	8
REDOVNOG	15	REDOVNOM	7	REDOVNOJ	5
RODNOG	31	RODNOM	37	RODNOJ	20
SELJAČKOG	4	SELJAČKOM	7	SELJAČKOJ	5
SEVERNOG	6	SEVERNOM	19	SEVERNOJ	13

Form -OG		Form -OM		Form -OJ	
Word form	Surface frequency (per 2 millions)	Word form	Surface frequency (per 2 millions)	Word form	Surface frequency (per 2 millions)
ŠIROKOG	17	ŠIROKOM	36	ŠIROKOJ	20
SJAJNOG	5	SJAJNOM	19	SJAJNOJ	10
SLAVNOG	15	SLAVNOM	6	SLAVNOJ	1
SLIČNOG	4	SLIČNOM	9	SLIČNOJ	4
SLOBODNOG	28	SLOBODNOM	17	SLOBODNOJ	29
SMRTNOG	11	SMRTNOM	8	SMRTNOJ	1
SPORTSKOG	10	SPORTSKOM	7	SPORTSKOJ	1
STAKLENOG	4	STAKLENOM	6	STAKLENOJ	2
STALNOG	44	STALNOM	22	STALNOJ	6
STRAŠNOG	23	STRAŠNOM	29	STRAŠNOJ	6
STRUČNOG	22	STRUČNOM	19	STRUČNOJ	5
ŠUMSKOG	14	ŠUMSKOM	13	ŠUMSKOJ	12
SUROVOG	7	SUROVOM	9	SUROVOJ	3
SVETSKOG	54	SVETSKOM	23	SVETSKOJ	15
TAČNOG	2	TAČNOM	7	TAČNOJ	2
TANKOG	17	TANKOM	25	TANKOJ	12
TEHNIČKOG	22	TEHNIČKOM	18	TEHNIČKOJ	13
TUŽNOG	34	TUŽNOM	16	TUŽNOJ	6
UMORNOG	28	UMORNOM	17	UMORNOJ	5
UPORNOG	2	UPORNOM	4	UPORNOJ	3
VAŽNOG	12	VAŽNOM	10	VAŽNOJ	2
VISOKOG	31	VISOKOM	29	VISOKOJ	25
VLAŽNOG	7	VLAŽNOM	22	VLAŽNOJ	7
VRELOG	18	VRELOM	26	VRELOJ	11
ŽALOSNOG	2	ŽALOSNOM	2	ŽALOSNOJ	2
ZLOG	6	ZLOM	9	ZLOJ	3
ZIMSKOG	12	ZIMSKOM	14	ZIMSKOJ	9
ŽIVOTNOG	15	ŽIVOTNOM	4	ŽIVOTNOJ	4
ZLATNOG	25	ZLATNOM	35	ZLATNOJ	12
ZNAČAJNOG	5	ZNAČAJNOM	3	ZNAČAJNOJ	2

Appendix B

Lemmas and their properties

			Relative Entropy				
			Case/number		Inflected forms		
Lemma	Lemma frequency (per 2 millions)	Lemma length (in letters)	Across gender	Feminine	Across gender	Feminine	Prepositon-adjectival phrases
atomski	105	7	0.210	0.168	0.159	0.086	1.719
beskrajan	303	9	0.038	0.034	0.069	0.024	0.768
blizak	248	6	0.071	0.116	0.105	0.046	0.711
bolestan	172	8	0.089	0.100	0.062	0.057	0.676
budan	182	5	0.123	0.137	0.123	0.074	3.039
buran	136	5	0.060	0.117	0.092	0.059	0.619
čeličan	94	7	0.068	0.094	0.119	0.014	0.970
čitav	478	5	0.050	0.123	0.086	0.097	1.200
čudan	407	5	0.038	0.026	0.052	0.012	0.649
čudesan	78	7	0.030	0.026	0.061	0.034	0.522
čvrst	232	5	0.035	0.090	0.059	0.058	0.481
davni	280	5	0.022	0.035	0.167	0.031	1.589
desni	125	5	0.100	0.177	0.135	0.075	1.241
dnevni	150	6	0.128	0.091	0.129	0.033	0.675
društven	288	8	0.083	0.098	0.100	0.034	0.389
dubok	646	5	0.033	0.035	0.072	0.049	0.796
ekonomski	333	9	0.074	0.085	0.111	0.024	0.532
filmski	136	7	0.113	0.152	0.170	0.054	0.417
glavni	418	6	0.076	0.043	0.171	0.012	0.737
gorak	362	5	0.024	0.027	0.049	0.031	0.713
gorski	115	6	0.063	0.107	0.117	0.040	2.350
gradski	618	7	0.061	0.054	0.106	0.036	0.495
gust	385	4	0.043	0.020	0.079	0.026	1.040
jak	528	3	0.104	0.166	0.119	0.079	0.552
južni	122	5	0.076	0.116	0.133	0.024	1.501
konačan	102	7	0.077	0.227	0.104	0.163	0.968
konkretan	77	9	0.135	0.173	0.077	0.071	0.400
kratak	404	6	0.060	0.062	0.039	0.015	0.567
kućni	150	5	0.134	0.101	0.254	0.031	0.801
kulturan	239	8	0.057	0.049	0.148	0.036	0.360
lak	471	3	0.065	0.068	0.113	0.073	1.347
lažan	98	5	0.053	0.045	0.079	0.047	0.315
lični	120	5	0.049	0.087	0.127	0.056	0.302
ljudski	567	7	0.053	0.011	0.088	0.011	1.273
malen	153	5	0.103	0.075	0.146	0.124	2.139

			Relative Entropy				
			Case/number		Inflected forms		
Lemma	Lemma frequency (per 2 millions)	Lemma length (in letters)	Across gender	Feminine	Across gender	Feminine	Prepositon-adjectival phrases
masovan	147	7	0.172	0.281	0.126	0.078	1.254
mek	522	3	0.037	0.044	0.084	0.021	0.896
mirisan	160	7	0.028	0.022	0.083	0.023	2.920
moćan	127	5	0.038	0.037	0.055	0.017	0.341
morski	174	6	0.057	0.019	0.099	0.014	0.628
mračan	294	6	0.030	0.026	0.030	0.002	0.426
mutan	481	5	0.023	0.043	0.037	0.007	1.335
muzički	151	7	0.035	0.112	0.191	0.029	0.358
naučni	103	6	0.053	0.120	0.164	0.090	0.274
nebeski	213	7	0.051	0.080	0.116	0.033	0.485
nečujan	101	7	0.089	0.116	0.110	0.057	4.646
nemiran	294	7	0.037	0.070	0.037	0.020	1.074
nepoznat	303	8	0.033	0.053	0.068	0.019	0.477
nevidljiv	211	9	0.016	0.029	0.025	0.005	0.619
nevin	154	5	0.041	0.050	0.034	0.013	0.720
nežan	313	5	0.037	0.033	0.056	0.021	0.602
neznan	227	6	0.051	0.054	0.103	0.011	1.974
nizak	239	5	0.036	0.046	0.047	0.011	0.763
noćan	289	5	0.030	0.049	0.121	0.025	0.482
običan	223	6	0.045	0.071	0.053	0.037	0.555
odličan	174	7	0.102	0.077	0.084	0.082	0.583
ogroman	345	7	0.030	0.019	0.030	0.013	0.704
otvoren	332	7	0.044	0.068	0.059	0.034	0.748
ozbiljan	166	8	0.074	0.097	0.072	0.063	0.425
plemenit	93	8	0.112	0.227	0.199	0.142	0.357
politički	488	9	0.048	0.058	0.140	0.020	0.379
poseban	173	7	0.072	0.059	0.048	0.032	1.253
povoljan	177	8	0.098	0.096	0.121	0.040	0.776
poznat	314	6	0.050	0.048	0.116	0.012	0.470
pravilan	95	8	0.166	0.107	0.087	0.069	1.135
prirodan	87	8	0.072	0.172	0.046	0.039	0.187
privredni	240	9	0.121	0.152	0.122	0.073	0.516
prljav	135	6	0.032	0.051	0.046	0.008	0.382
proklet	99	7	0.068	0.108	0.086	0.069	1.046
radostan	212	8	0.057	0.045	0.063	0.019	1.947
ranjen	234	6	0.042	0.066	0.048	0.037	1.344
ratni	275	5	0.094	0.088	0.158	0.068	1.121
redovan	110	7	0.055	0.158	0.060	0.051	0.392
rodan	330	5	0.036	0.047	0.140	0.043	1.321

			Relative Entropy				
			Case/number		Inflected forms		
Lemma	Lemma frequency (per 2 millions)	Lemma length (in letters)	Across gender	Feminine	Across gender	Feminine	Prepositon-adjectival phrases
seljački	140	8	0.099	0.081	0.133	0.022	1.093
severni	119	7	0.179	0.257	0.212	0.107	1.207
širok	688	5	0.030	0.023	0.057	0.007	0.327
sjajan	276	6	0.038	0.036	0.076	0.006	0.656
slavan	136	6	0.046	0.065	0.063	0.012	0.801
sličan	344	6	0.065	0.064	0.091	0.014	0.398
slobodan	494	8	0.030	0.075	0.036	0.017	0.420
smrtan	136	6	0.081	0.066	0.090	0.028	0.581
sportski	138	8	0.045	0.052	0.161	0.028	0.437
staklen	101	7	0.050	0.072	0.053	0.010	1.024
stalan	227	6	0.059	0.056	0.064	0.031	0.789
strašan	318	7	0.022	0.023	0.100	0.034	0.656
stručan	230	7	0.079	0.076	0.107	0.044	0.271
šumski	199	6	0.071	0.089	0.136	0.046	0.841
surov	103	5	0.045	0.045	0.048	0.018	0.350
svetski	258	7	0.121	0.236	0.134	0.052	0.492
tačan	121	5	0.196	0.092	0.197	0.038	1.309
tanak	394	5	0.064	0.059	0.068	0.020	0.736
tehnički	218	8	0.066	0.098	0.085	0.010	0.404
tužan	689	5	0.050	0.052	0.067	0.033	0.701
umoran	602	6	0.069	0.106	0.072	0.026	1.932
uporan	79	6	0.066	0.097	0.124	0.035	3.983
važan	334	5	0.061	0.086	0.060	0.018	3.318
visok	738	5	0.025	0.034	0.054	0.013	2.921
vlažan	201	6	0.042	0.019	0.049	0.013	3.524
vreo	431	4	0.035	0.023	0.054	0.015	3.499
žalostan	115	8	0.092	0.075	0.134	0.078	3.817
zao	195	3	0.043	0.045	0.068	0.018	4.505
zimski	194	6	0.053	0.046	0.178	0.016	3.209
životan	82	7	0.073	0.088	0.118	0.013	4.590
zlatan	566	6	0.037	0.039	0.042	0.004	3.109
značajan	205	8	0.077	0.136	0.072	0.035	3.030