

 Open access • Journal Article • DOI:10.1556/ALING.58.2011.1-2.4

Allomorphic responses in Serbian pseudo-nouns as a result of analogical learning

— [Source link](#) 

Petar Milin, Emmanuel Keuleers, Dušica Filipović Đurđević

Institutions: University of Novi Sad

Published on: 09 Mar 2011 - Acta Linguistica Hungarica (Akadémiai Kiadó)

Related papers:

- [Analogical Levelling and Optimisation: The Treatment of Pointless Lexical Allomorphy in Greek](#)
- [Testing the Semantic Homogeneity Constraint: Analogical change and Russian verbs](#)
- [Rules vs. analogy in English past tenses: a computational/experimental study](#)
- [Learning biases predict a word order universal](#)
- [On phonically based analogy](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/allomorphic-responses-in-serbian-pseudo-nouns-as-a-result-of-i2we30tk8>

Allomorphic responses in Serbian pseudo-nouns as a result of analogical learning

Petar Milin^{a, c}, Emmanuel Keuleers^b, Dušica Filipović Đurđević^{a, c}

^a Department of Psychology, University of Novi Sad, Serbia

^b Department of Experimental Psychology, Ghent University, Belgium

^c Laboratory for Experimental Psychology, University of Belgrade, Serbia

Abstract: Allomorphy is a phenomenon that occurs in many languages. Several psycholinguistic studies have shown that allomorphy, if present, co-determines cognitive processing. In the present paper we discussed allomorphic variations of Serbian instrumental singular form of pseudo-nouns as emerging from analogical learning. We compared the predictions derived from memory-based language processing models with results from previous experimental study with adult Serbian native speakers. Results confirmed that production of suffix allomorphs in Serbian instrumental singular masculine nouns could be accounted for by memory-based learning, and simple analogical inferences. The present findings are in line with a growing body of research showing that memory-based learning models make relevant predictions about the cognitive processes involved in various linguistic phenomena.

Keywords: allomorphy, memory-based learning, analogy, Wug-task, Serbian

Introduction

In this paper we will present a probabilistic computational model of allomorphy and demonstrate that allomorphic variation may arise from analogical learning of the mapping from stems to inflected forms. We will make use of behavioral experiments that were previously conducted with adult native speakers of Serbian engaged in a computerized *Wug* task (Jovanović, 2008; see Berko, 1958 for the original *Wug* task experiment). Looking at the two allomorphic forms of the instrumental singular of Serbian masculine pseudo-nouns, we will compare the performance of native speakers with the outcomes of several simulations using computational models of analogical learning.

The allomorphy represents a variation in the form of a particular morpheme, without a change in its meaning (cf. Lieber, 1982; Lyons, 1986; Spencer, 2001 etc.). In English, variations in the *-ed* morpheme used in the regular past tense, and the *-s* morpheme used to mark noun plurals, are well known examples. The regular English past tense suffix appears in three different forms (or *morphs*), depending on the final sound of the verbal stem: walk-ed (/t/), jogg-ed (/d/), trott-ed (/əd/). In modern Arabic, allomorphy occurs in an *etymon* – a bi-consonantal morphological unit that carries semantic information of a given word (Ratcliffe, 1998; Boudelaa & Marslen-Wilson, 2001; Boudelaa & Marslen-Wilson, 2004). In Dutch, the diminutive suffix has two frequent allomorphic variations (*-tje* and *-je*), and three less frequent ones (*-etje*, *-pje* and *-kje*) (Daelemans, Berck & Gillis, 1997). In Finnish, allomorphy appears both in the stem (Järvikivi & Niemi, 2002) and in suffixes (Järvikivi, Bertram & Niemi, 2006). Similarly, in Hungarian, allomorphic variations occur as stem shortening or lengthening (Pléh, Lukács & Racsmány, 2002), and as suffixal alternating vowels (Kertész, 2003; Hayes & Cziráky-Londe, 2006). Finally, allomorphy is present in Slavic languages as well. Affixal allomorphy in Russian is discussed in detail by Blevins (2004), while Ivić (1990) and Zec (2006) provided linguistic analysis of the suffix allomorphy in Serbian instrumental singular masculine and neuter nouns.

Allomorphy as a cognitive phenomenon

For cognitive science, and in particular for psycholinguistics, the main question of any language phenomenon is its cognitive relevance. If a particular linguistic phenomenon can also produce critical differences in behavioral and/or neurological measures, then one can say that the linguistic phenomenon also has cognitive relevance. Although often not of central interest, the cognitive relevance of allomorphy has repeatedly been attested in behavioral research. Schreuder and Baayen (1995) stated that we may be slower in processing words with affixes that have several allomorphs, than words containing affixes for which there is no allomorphic variation. Järvikivi, Bertram & Niemi (2006) made a similar, but more detailed claim, using the concept of *affixal salience* – "the probability with which an affix is likely to emerge from the orthographic/phonological string" (p. 395). They showed that affixal salience decreases as the number of affixal allomorphs increases. Conversely, however, to the inhibition that allomorphy produced to a

given affix, allomorphic realizations of bound nominal stems in Finnish significantly primed the same noun in its base form – nominative singular (Järvikivi & Niemi, 2002a; Järvikivi & Niemi, 2002b). Similarly, in a priming task in Spanish, Allen & Badecker (1999) found no difference between conditions in which the prime was a true stem-homograph of the target (e.g., "placer" (*pleasure, to please/inf./*) – "placa" (*plate, panel*)) and conditions in which the target was preceded by a stem allomorph of the prime (e.g., "plazca" (*to please/subjunctive 3 Pers. Sg./*) – "placa" (*plate, panel*)). Finally, specific difficulties in processing allomorphic variations in Hungarian nouns were observed with normal children (Pléh, 1989), and with children with Williams syndrome (Pléh, Lukács, & Racsmány, 2002).

One of the most common instances of allomorphy in Serbian is the suffix allomorphy (*-em* vs. *-om*) occurring for instrumental (*making use of*) singular masculine nouns. For instance, Serbian native speakers may be somewhat puzzled whether to say "nos-*om*" or "nos-*em*" (*using the nose*), "malj-*om*" or "malj-*em*" (*using an odor*), "obruč-*om*" or "obruč-*em*" (*using a hoop*), "pištolj-*om*" or "pištolj-*em*" (*using a revolver*), and so on. Jovanović et al. (2008) directly addressed this form of allomorphy using two experimental tasks. First, using a *sentence completion task*, the authors confirmed that suffix allomorphy in Serbian instrumental singular masculine nouns occurred only when a noun stem ended in a particular subset of consonants: *palato-alveolars* or *back coronals*.¹ Second, using a visual lexical decision task, they showed that suffix allomorphy in Serbian masculine nouns, with stem ending in back coronals, elicits significant differences in processing latencies: for words with the *-om* suffix, an increase in observed form frequency was associated with an increase in processing latency, while for the *-em* suffix, an increase in form frequency was associated with a decrease in reaction time. This interaction between a particular suffix realization (*-om* or *-em*) and its probability in production task showed that even though *-om* is the most frequent suffix in the Serbian instrumental singular, it is processed slower if encountered within the phonological domain for which *-em* is preferred. Although such complementarity

¹ Different subset labels come from two means of consonant classification. Front coronals match alveolars: **n** (/n/), **l** (/l/) and **r** (/r/) and include five additional consonants: **t** (/t/), **d** (/d/), **s** (/s/), **z** (/z/) and **c** (/tʃ/). Back coronals match palato-alveolars: **č** (/tʃ/), **ć** (/tɕ/), **dž** (/dʒ/), **đ** (/dʒ̞/), **nj** (/ɲ/), **lj** (/ʎ/), **j** (/j/), **š** (/ʃ/) and **ž** (/ʒ/).

might suggest rule-based derivation of the two allomorphic forms, we will advocate that this pattern can emerge from a more parsimonious learning principle.

Modeling allomorphic response as analogical learning

Jovanović and her collaborators (2008) discussed their results in respect to previous findings of Mirković, Seidenberg & Joanisse (2009), who used a connectionist network to model the production of Serbian case-inflected morphology. This model used a training set of 3244 Serbian nouns, and learned to produce the correct case-endings by developing particular probabilistic constraints at the level of phonology and semantics. At the end of learning, the error rate for masculine instrumental singular – our target case, was still approximately 4%. However, the model excluded the possibility of having both suffixes applied to the same stem with different probabilities, but implemented a simple rule that attached either *-om* or *-em* to a given stem. For instance, all masculine nouns with a stem ending in an alveolar or palato-alveolar consonant, necessarily took the *-em* suffix, while all nouns with other terminating consonants used *-om* instead (Mirković et al., 2009).

In contrast, a study by Zec (2005) showed that masculine noun stems ending with a coronal can, and usually do have allomorphic realizations in the instrumental singular: both *-om* and *-em* can apply. Jovanović and colleagues (2008) and Jovanović (2008) confirmed the analysis of Zec (2005), both in a lexical decision task and in a computerized modification of the *Wug* task (Berko, 1958), administered to adult native speakers of Serbian. More specifically, masculine nouns ending in back coronals (or palato-alveolars) were significantly more likely to allow for both suffix allomorphs (*-om* and *-em*).

In principle, connectionist networks should be capable of modeling allomorphy. In particular, a probabilistic version of the model of Mirković and collaborators (2009) could account for allomorphic variation in Serbian nouns. However, the immense power or flexibility that is typical for artificial neural networks, comes at a cost of lacking insight in how a given network achieved a particular morphological mapping. As Norris (2005) suggested, the true contribution of connectionist models should not come from their performance, but from understanding the principles that guide the performance of the networks (see also Baayen, 2003 for a more elaborate discussion). Thus, the question is whether more directly addressable learning

mechanisms could meet the same goal. In particular, we are interested in testing whether we could model allomorphic variation in Serbian instrumental singular by using a very simple analogical approach. However, before we go any further, a note of caution is in order: it is perfectly possible to successfully model the same phenomenon using different *machine learning* approaches. What is important is the contribution that different approaches give to our understanding of the phenomenon. Following Marr (1982), we can say that analogical learning improves our understanding mostly at the *algorithmic level*, revealing the processes and representations of this task. At the same time, a connectionist network improves our understanding mostly at the *implementational level*, showing how neural structures and neuronal activities might implement a given cognitive task.

Our claim is that allomorphy can take place from analogical inference, where sources of analogy (existing stem forms) compete with each other in providing one or the other suffix allomorph – possible inflected forms of instrumental singular masculine nouns. Acquisition and processing of linguistic knowledge by means of memory and analogy has a long history in twentieth-century linguistics (De Saussure, 1916; Bloomfield, 1933; Harris, 1951; 1957 etc.). Recently, the idea has been further developed by usage-based models of language (Bybee, 2007). In psychology, the concept of analogy can be seen in exemplar-based accounts of human categorization behavior (Smith & Medin, 1981; Nosofsky, 1986; Estes, 1994). According to these accounts, categories are formed by storing exemplars in memory, and categorization decisions are made by relying on similarities of target stimuli to exemplars stored in memory. In computational linguistics, these ideas have been applied in memory-based learning (Daelemans & Van den Bosch, 2005) and Analogical Modeling of Language (Skousen, 2002).

According to the memory-based learning approach, a categorization decision (e.g., the choice of allomorph) is resolved by re-use of existing exemplars and analogical reasoning. In order for this process to take place, three conditions need to be fulfilled. In the case we are studying here, firstly, we need a store of exemplars (stems) with assigned exponent (the instrumental ending). These exemplars can be represented as vectors of phonological features at the subsyllabic level (i.e., the onset, nucleus, coda of each syllable). Secondly, a distance function is required to compute the similarity of the target form to the forms stored in memory. Finally, in order to assign a class to the novel exemplar, a decision function is required. The

decision function is adopted from the field of artificial intelligence and is based on the k nearest neighbor classifier method (k -NN). This implies that the outcome of the decision function is determined by the class of the k nearest neighbors (e.g., if $k = 1$, a novel exemplar is assigned a class of the exemplar most similar to it). Memory-based learning has a long history of application within the field of computational linguistics. Recently, the method has also been successfully applied in psycholinguistic research, where the aim is to approach the performance of native speakers, that is, to simulate the functioning of the cognitive system. By now, a considerable body of empirical data demonstrated the efficiency of memory-based learning. Keuleers et al. (2007) and Keuleers and Daelemans (2007) have demonstrated that outcomes of simulations based on the memory-based learning paradigm mimic performance of native speakers in the production of Dutch noun plurals. Similar findings have been reported for Italian verb conjugations (Eddington, 2002a), Spanish gender assignment (Eddington, 2002b), linking elements in German compounds (Krott, Schreuder, Baayen and Dressler, 2007) and so on.

Problem

In this paper, we will compare the predictions derived from memory-based learning models to experimental results by looking at production of allomorphic variations using pseudo-nouns in the domain of the Serbian instrumental singular. Attempts have been made in describing orthographic/phonological properties of stems that lead to the production of each of the two allomorphic variations (Zec, 2006 in particular). These descriptions were moderately successful in predicting responses collected from native speakers, and can be seen as rules for choosing an allomorph. In this study, we will not compare the predictions derived from these rules to the results obtained by means of exemplar-based modeling. Our aim is to demonstrate that analogical learning can account for allomorphic variation at least as well as the rule-based descriptions. Moreover, the difference between the analogical models and the rule-based descriptions is that the former operate in a completely inductive manner. By this we imply that the model does not rely on *a priori* knowledge of which features are important and which ones are not.

The predictive power of the memory-based learning models will be tested by comparing the outcomes of simulations to behavioral responses collected from

native speakers. In particular, for each allomorph, we will be looking at the correlation coefficients between probabilities assigned by the model and the probabilities observed in behavior of native speakers (by dividing speakers preferring one allomorph with total number of speakers in a given sample). Because the simulations are based on the principles of memory-based learning, high correlation coefficients between these probabilities would suggest that these principles have a cognitive relevance.

Finally, the memory-based learning models will use only similarity between forms at the level of orthography/phonology.² Although a clear improvement in predictions is to be expected if additional similarities were included, we shall opt for simplicity, and examine the explanatory potential of a simple measure.

Method

Experimental data

The experimental data are taken from Jovanović (2008). In total, 42 adult participants, first year students of Psychology in Novi Sad, mainly females, with normal or corrected-to-normal vision participated in a computerized Wug-task. Jovanović used 125 pseudo-stems that followed Serbian ortho-phonotactic constraints. Each pseudo-stem was exactly five characters long, and had a fixed CVCVC structure. The final VC segment was controlled: all 25 Serbian consonants occurred five times as a final consonant, preceded once with each of the five vocals. For example, some of the pseudo-stems used in experiment were: "bobaš", "cogilj", "gofić", "nirib", "salav" and so on. To implement the Wug-task, Jovanović downloaded 125 pictures from the *What is it?* web-site (<http://puzzlephotos.blogspot.com>). Each trial started with presentation of an unknown picture with its pseudo-word label in nominative singular (for 2000 ms). Then, a grammatical Serbian sentence appeared with the critical pseudo-word in both of instrumental singular allomorphs. One allomorph was positioned a row above and the other was positioned a row below blank space that was in line with the rest

² Serbian has shallow orthography, and mapping from phonology to orthography is one to one. Hence, for the purpose of present research, this difference can be disregarded.

of the words forming a sentence (for example: "Motori se testiraju cogiljem/cogiljom."; in English: "Engines are tested (by) cogiljem/cogiljom."). The participants' task was to choose one of the two forms by pressing a spatially corresponding button. There was no response time-out. It took approximately 10 minutes for participant to complete the task. Based on participants' choice, probability of each of the two allomorphic forms was estimated.

Simulation procedure

Implementation of the memory-based learning model started with the selection of an exemplar-storage that made up the "memory" of the model. For the present research, we used all 3481 masculine and neuter nouns from the *Frequency Dictionary of Contemporary Serbian Language* (Kostić, 1999), which occurred in instrumental singular case. Neuter nouns were included because their instrumental singular can also attach both *-om* and *-em* suffix, depending on the final vowel (*-o* or *-e*). This inclusion gave additional noise in the exemplar-storage, thus making analogical learning more demanding.

In memory-based learning, the problem of predicting an allomorph is considered a simple classification problem: each pseudo-word needs to be classified as taking *-om* or *-em*. For this, the memory base was searched for the k nearest neighbors. For instance, in a model where the neighborhood size (k) equals 7, we would search the memory for the 7 stems that were most similar to the pseudo-word.³ We could then look at how often the *-em* and *-om* suffixes occurred among these stems. The estimated probability of each suffix then was a simple ratio of the times it occurred in the neighborhood to the total number of stems in that neighborhood. We tested models with different neighborhood sizes: we linearly increased k from 1 to 16, after which we used an exponential growth function of base 2 ($k = 32, 64, \dots, 1024, 2048$), until finally k equalled the size of the lexicon (3481 items).

In addition to the parameter k , memory-based learning models have another two crucial parameters: the *distance metric* used for computing the similarity between

³ In practice, the parameter k refers to nearest distances rather than nearest neighbors. When several exemplars occur at the same distance from the target, these exemplars are considered tied. In other words, a k -NN model looks at least k exemplars. See Keuleers and Daelemans (2007) for a more detailed treatment of this issue.

exemplars stored in memory and the pseudo-word to be classified, and the *decay function*, defining how a neighbor's weight in the classification decreases with distance from the target pseudo-word. We employed three well-known distance metrics: Jeffrey divergence, Levenshtein distance, and Hamming or Overlap distance. The Overlap metric is the coarsest: it simply counts the number of mismatching features. The Levenshtein distance is a generalized version of Overlap distance: it measures how many features must be inserted, deleted, or replaced to transform the stem into the pseudoword. Finally, Jeffrey divergence uses principles from information theory to give a weight to each feature, and operates as a weighted Overlap metric (for an in depth presentation of these measures consult Rubner, Tomasi & Guibas, 2000; Levenshtein, 1966; Hamming, 1950; for their application in linguistics see Daelemans & Van den Bosch, 2005). In addition to the distance metrics, we compared three decay functions: Zero Decay, where all neighbors have the same influence on classification, regardless of their distance to the pseudo-word; Inverse Distance Decay, where neighbors are weighted by the inverse of their distance; and Exponential Decay, where a neighbor's weight decreases exponentially with its distance. Since both the neighborhood size, the definition of similarity and its decay weighting affect the composition of the neighborhood, these parameters can interactively affect the outcome of a simulation.

Results

In the very first step of analysis we estimated the probability of producing *-om* and *-em* suffix for each noun based on participants' responses in Wug-task. These probabilities were then correlated with the outcomes of the memory-based learning simulations, where *distance metric*, *decay weight* and *neighborhood size* were systematically varied as factors. These results are presented in Figure 1.

As we can see from the plots, the similarity between human and computer results, expressed in terms of product-moment correlation coefficient, reached its maximum very rapidly. This means that in most cases, a very small number of exemplars was sufficient for memory-based learning to make a correct analogy and to produce human-like output of suffix allomorphy in Serbian instrumental singular pseudo-nouns. After including about ten nearest neighbors, nothing much could be gained, as the right-hand lines presenting exponential increase of neighbors show.

Moreover, without decay weighting any further increase in number of neighbors was harmful for the similarity between human and computer-simulated behavior, while exponential and inverse decay weights just alleviated cost of using large neighborhoods.

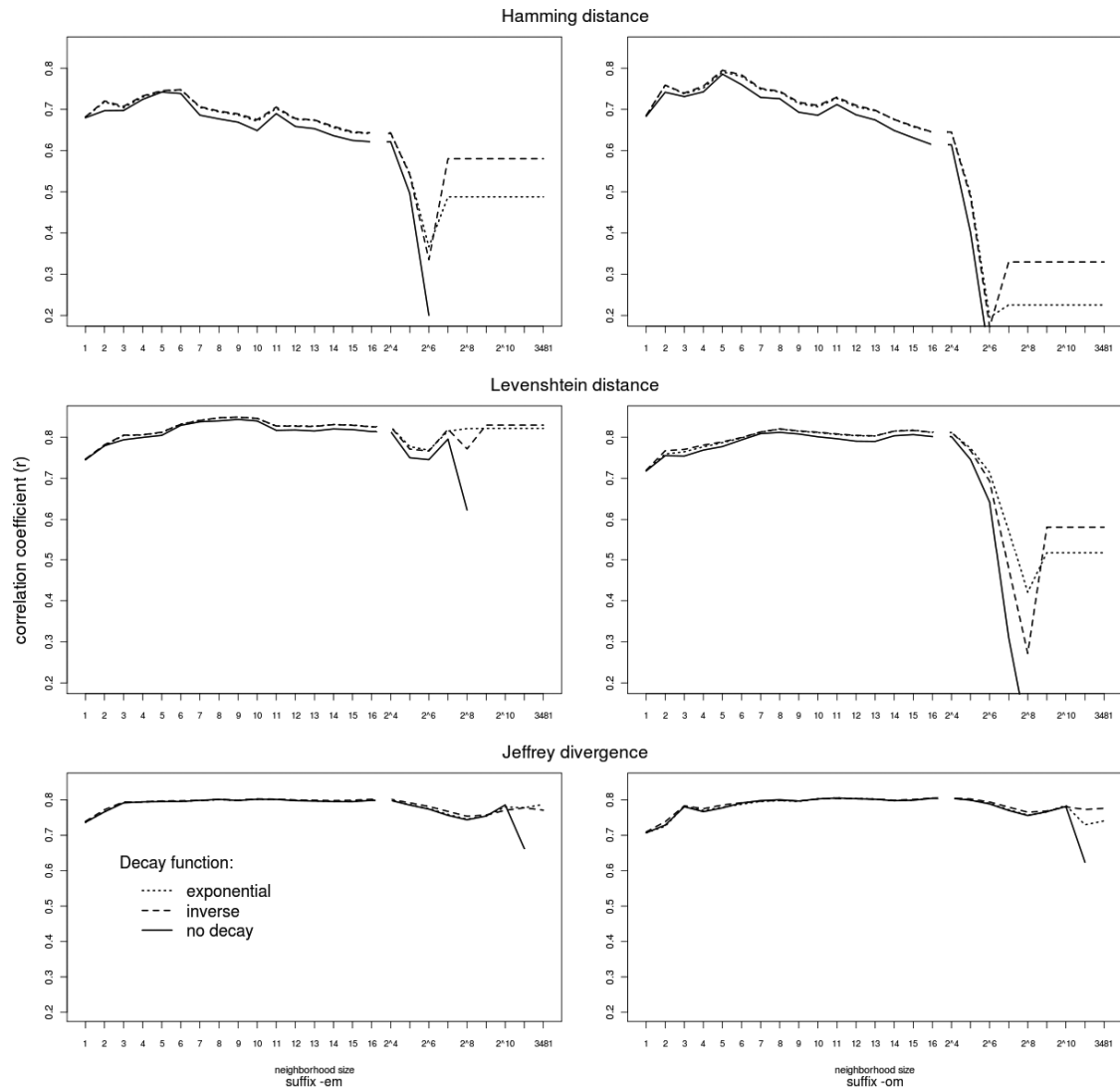


Figure 1. Correlation coefficients between probabilities of producing -om and -em suffix allomorph, in behavioral experiment (Wug-task) and computer simulations. Line-breaks mark points where increase in neighborhood size changes from linear to exponential.

Row-wise comparisons of graphs by means of visual inspection already show that there were no substantial differences between the three distance metrics. However, in addition to visual inspection of graphs, we performed more detailed statistical comparisons of the three distance measures. Having two allomorphs crossed with three distance measures and three decay weights for each number of neighbors provided us with a total of 18 correlation coefficients per number of neighbors. In

order to demonstrate that there were no significant differences between 18 correlation coefficients within a given number of neighbors, we tested for the significance of the difference between the smallest and the largest correlation coefficient for each of the first sixteen neighborhood sizes, separately. If the difference between the smallest and the largest of correlation coefficients was not significant, then we could deduce that none of the differences were. In other words, this way we could demonstrate that all three distance measures using three different decay weighting performed equally well both for *-em* and for *-om* forms, for a given number of neighbors. The tests confirmed the null-hypothesis, thus proving that, in range from one to sixteen nearest neighbors, with any of the three measures using any of the three decay weighting we can achieve approximately the same success in simulating human production.

However, some variations were rather interesting and specific to a given measure. Using the simplest of the three measures – Hamming's distance (i.e., Overlap), gave somewhat lower correlations, but Jeffrey divergence, although the most sophisticated measure, did not perform better than Levenshtein distance. However, using Jeffrey divergence, the difference in similarity in producing each of the two allomorphic variants (ending with *-om* or with *-em*) was negligible. Levenshtein distance provided a better match to the human responses for the *-em* allomorph, while the Hamming distance did exactly the opposite. Finally, larger neighborhoods were the least penalizing for Jeffrey divergence. It seems that this was the single point where some leverage from a more sophisticated measure was observed. This finding might be surprising, but can be simply explained by the fact that Jeffrey divergence is more fine-grained than the other measures. It expresses distances as real numbers. Therefore exemplars do not tie often at the same distance, while the Overlap and Levenshtein metric, which express distances in integer numbers, collapse many exemplars at the same distance. Jeffrey divergence reaches the same neighborhood size in absolute terms (the total number of exemplars) at a much later point than the other similarity metrics, thus having particular decay weighting as its intrinsic property.

In order to make comparison of human and simulated behavior even more rigorous and conservative, we developed a specific statistical procedure which made use of logistic mixed-effect regression modeling (c.f., Baayen, Davidson & Bates, 2008; Jaeger, 2008 etc.). Firstly, we ran our analysis for each distance metric and each

decay weighting, separately. Secondly, for a given distance metric (Jeffrey, Levenshtein, Hamming) with a given decay weighting (no decay, exponential decay, inverse decay), we iteratively applied linear-mixed modeling to test for predictability of a particular number of neighbors used in memory-based learning simulation run. We tested range of neighbors from one to sixteen. Each step in the procedure compared two statistical models: the *more specific* model, which included probabilities from simulation with *more general* model, which included those probabilities and, additionally, residual probabilities of $k+1$ neighbors taking out probabilities of k neighbors. In other words, we included only *novel* variability in probabilities from the simulation with $k+1$ neighbors, the one that was not already present in the probabilities from k neighbors. The statistical models used the binomially distributed participants' response – selecting *-om* or *-em* suffix allomorph, as a dependent variable, while items (pseudo-words) and participants were treated as random-effects. Two successive models were compared by applying likelihood-ratio tests, which produced Chi-squared values and corresponding p -values that are listed in Table 1.

k	HAMMING DISTANCE			LEVENSHTEIN DISTANCE			JEFFREY DIVERGENCE		
	no decay	exp. decay	inv. decay	no decay	exp. decay	inv. decay	no decay	exp. decay	inv. decay
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.003	0.001	0.001	0.000	0.001	0.000	0.002	0.002	0.001
2	0.141	0.310	0.312	0.036	0.004	0.005	0.001	0.001	0.001
3	0.008	0.009	0.009	0.036	0.108	0.109	0.157	0.236	0.363
4	0.015	0.030	0.039	0.044	0.034	0.034	0.639	0.446	0.366
5	0.044	0.033	0.038	0.001	0.002	0.003	0.661	0.781	0.714
6	0.468	0.683	0.727	0.027	0.030	0.032	0.393	0.501	0.719
7	0.769	0.919	0.949	0.062	0.035	0.023	0.146	0.155	0.138
8	0.193	0.262	0.243	0.143	0.210	0.249	0.755	0.676	0.836
9	0.770	0.737	0.734	0.788	0.948	0.976	0.128	0.131	0.107
10	0.003	0.009	0.009	0.221	0.205	0.218	0.863	0.767	0.947
11	0.646	0.635	0.600	0.365	0.399	0.556	0.378	0.346	0.384
12	0.219	0.204	0.254	0.949	0.960	0.738	0.651	0.679	0.743
13	0.462	0.643	0.565	0.086	0.113	0.101	0.865	0.688	0.853
14	0.344	0.472	0.477	0.871	0.734	0.687	0.757	0.512	0.309
15	0.678	0.698	0.736	0.785	0.908	0.822	0.125	0.125	0.167

Table 1. The likelihood-ratio test p -values for a series of successive mixed-effect models with k and $k+1$ residual probabilities as covariate predictors of participants' binary response. Significant p -value ($p < 0.05$) reads as significant contribution of $k+1$ residuals.

The first line in the table can be read as follows: *Were the probabilities obtained from a memory-based learning model with a neighborhood size of one a significant predictor of the participants' choice of allomorph?* Given the very small p -values in the columns corresponding to each metric, the answer to the question ought to be that a memory-based learning model with a single nearest neighbor was a very significant predictor of the participants' choice of allomorph, regardless of the similarity metric and decay weighting used. The second line, and all subsequent lines, should be interpreted in the following way: *Did increasing the neighborhood size by one significantly improve the prediction of the memory-based learning model compared to the model with the previous neighborhood size?* The Hamming distance, which is the coarsest measure of all three, demonstrated somewhat shivering predictiveness; increasing the neighborhood size might increase or decrease model fit in first few steps, with a sudden upraise for the model with a neighborhood size of eleven. The Levenshtein metric had a much more regular behavior. At each step from one to seven, the model significantly gained in fit. If we recall that the Levenshtein distance uses only simple operations of insertion, deletion and substitution of feature values when expressing the distance between two exemplars, then it is striking that using this metric made for better predictions than using Jeffrey's divergence, a more complex measure with a higher information load. With Jeffrey divergence, increasing the neighborhood size improved the fit at each step, until a model with a neighborhood size of three was reached. However, Jeffrey's metric with three neighbors and without decay weighting gave a slightly worse fit than the model with a Levenshtein metric with seven neighbors and no decay weighting. This is indicated both by smaller beta estimate and z -statistic ($\beta = 3.267$, $z = 11.82$, $p < 0.001$ and $\beta = 4.041$, $z = 13.74$, $p < 0.001$), and measures of goodness-of-fit (greater AIC: 4612 vs. 4590; and smaller log-likelihood value: -2302 vs. -2291). All this leads to a conclusion that Levenshtein distance is not just a simpler, but also a better predictor of human responses. It only needs a larger neighborhood for proper analogical inferences.

After closer examination of the candidate model (using Levenshtein distance, no decay weighting and with seven nearest neighbors), another interesting pattern of results was revealed. Our critical predictor covariate – probabilities from simulation using Levenshtein distances with $k = 7$ nearest neighbors, entered into significant interaction with random-effect of participants. In other words, in addition to the by-

participant adjustment for the intercept, we needed to let loose the slope of the predictor allowing it to vary across participants. Furthermore, results showed significant correlation between by-participant random intercepts and by-participant random slopes for simulated probabilities ($r = -.60$, after removing spurious residuals). Although this might seem discouraging for interpretation and understanding, it actually uncovered detailed interplay between probabilities obtained from computer simulation run and that from participants' responses. Firstly, adjustments for intercept show that participants differ significantly in their "readiness" to produce *-em*-ending variant. Secondly, although, as expected, we observed significant positive correlation between simulated and human probabilities in producing variant with the suffix *-em*, this correlation varied across participants: changes in probabilities for the *-em* suffix variant matched changes obtained in computer simulation more tightly for some participants than others. Finally, strong negative correlation between by-participant random intercepts and by-participant random slopes for simulated probabilities told us that the higher the base probability for producing *-em* variant for a given participant, the less tightly she/he matched changes in computer simulated probabilities. However, one should keep in mind that the observed variation did not affect the overall predictability of simulation outcomes, it only revealed additional peculiarities in participants' behavior.

By-participant variations in the intercept and the slope for simulated probabilities are presented on Figure 2. As we can ascertain from the left panel, there is balanced number of participants that produce *-em*, as well as those who produce *-om* with the higher odds. Similarly, adjustments for the slope, represented on the right panel, are scattered on the both sides as well. It came out that computer simulated probabilities appear as *prototypical* or average participant, while real participants vary around. And this was to be expected too: analogical learning had taken place from the *Frequency Dictionary of Contemporary Serbian Language* (Kostić, 1999), as exemplar-storage, where allomorphic variants for masculine and neuter instrumental singular were averaged across many native speakers. Thus, exemplars from multitude became wide and middling, leading to the analogical inferences of a typical native speaker of Serbian language.

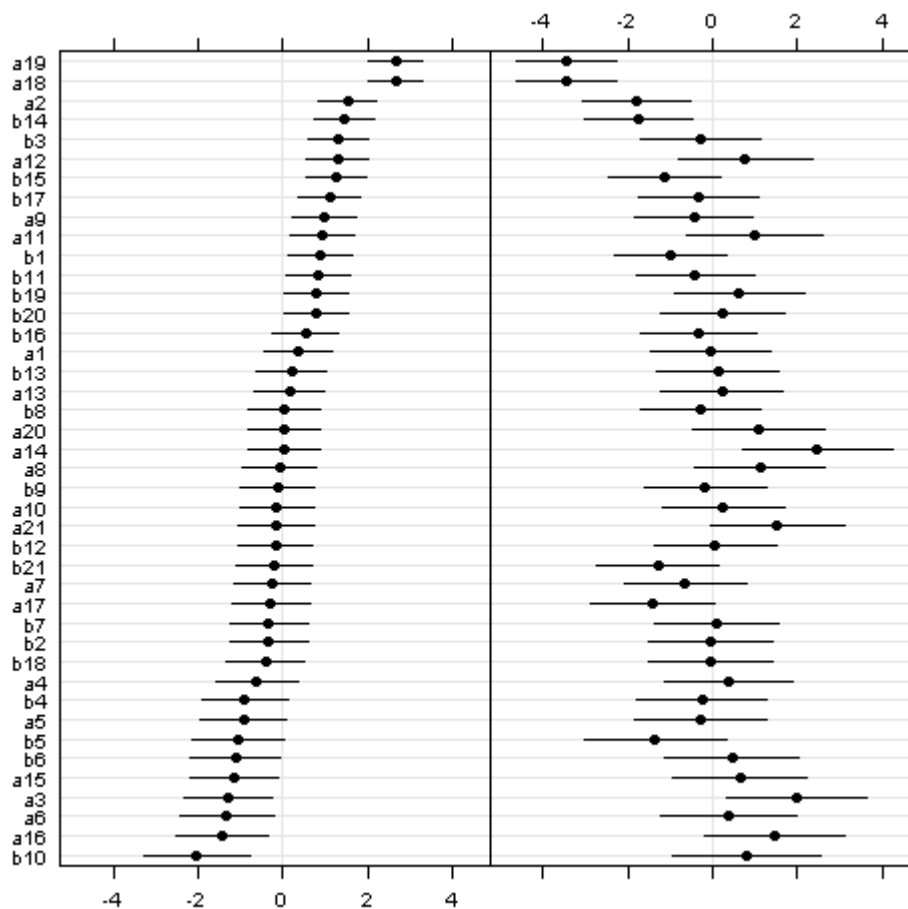


Figure 2. Visualization of the by-participants adjustments for the intercept (left-hand panel) and slope (right-hand panel) of the probabilities obtained in computer simulation using Levenshtein distance with $k=7$ nearest neighbors. Both panels are centered to the grand intercept and grand slope, thus, values greater than zero correspond to higher intercept value and steeper slope for a given participant, and vice versa.

Discussion

We aimed at demonstrating that production of allomorphs of instrumental singular of Serbian masculine nouns can be accounted for by memory-based learning. We started by implementing a memory-based learning model of the allomorphy in question and comparing probabilities assigned by the model with probabilities of each of the allomorphs being produced by native speakers. Our analyses showed that outcomes of the model closely resembled native speakers' behavior. We make no attempt in claiming that the model architecture mirrors the organization and processing performed by cognitive system. However, we wish to state that in

principle, the patterns observed in the behavior of native speakers can be accounted for by a very simple learning principle. This would argue against application of rules in describing linguistic phenomena. Interestingly, predictions of the model were equally successful for *-em* and *-om* suffixes, allowing our model to argue against "default" accounts of language.

In the model we applied, the probability of a given allomorphic form was a result of the analogical inference based on a simple orthographic similarity between the stem and a small number of exemplars stored in the memory. We compared predictions derived from three measures of orthographic similarity. Our results showed that the three measures were highly similar in predicting human responses. For each of the measures, strong resemblance to native speakers' behavior could be achieved even when analogical inference was based on only one exemplar (the one that is most similar to the stem in question). As expected, resemblance to native speakers' behavior increased with an increase in the number of exemplars evoked from the memory, but only up to a certain neighborhood size. Moreover, at a certain point, taking more exemplars into account degraded the performance of the model. The more exemplars were taken into account, the more a model resembled a frequency based approach: if all exemplars in the model's memory are taken into account, the output of the model is simply the ratio of *-em* vs. *-om* suffix frequencies in the memory. This suggests that making decisions independent of similarity would be a bad strategy. Although the three similarity measures were equally successful as predictors, the speed of the observed degradation (produced by an increase in neighborhood size) differed. Additionally, introduction of exponential or inverse decay weighting appeared as beneficial in larger and large neighborhoods, diminishing overall degradation.

Detailed analyses that took into account both simplicity of similarity measures and the speed of degradation that followed increase in neighborhood size, demonstrated that the optimal solution uses a simple measure and a neighborhood of seven exemplars. From this conclusion follows almost indecent question whether this number could be "magical", not only for working memory load (Miller, 1957), but for language processing, as well. Unfortunately, present findings from one language and concerning one specific phenomenon cannot provide the answer.

The observed results add to a growing body of research showing that memory-based learning models make relevant predictions about the cognitive processes involved in

various linguistic phenomena, such as formation of Dutch plurals (Keuleers et al., 2007; Keuleers & Daelemans, 2007), Dutch diminutives (Daelemans, Berck, & Gillis, 1997), English past tense (Keuleers, 2008), German plurals (Hahn & Nakisa, 2000), Spanish gender assignment (Eddington, 2002b) and so on. On a more general level, these findings fit well with the *gradient* view of various linguistic phenomena. This framework solicits for continual as opposed to discrete transitions between linguistic categories (cf., Albright & Hayes, 2003; Hay & Baayen, 2005; Baayen, Fledman, & Schreuder, 2006; Bybee, 2007; Keuleers et al., 2007; Milin, Filipović Đurđević, & Moscoso del Prado Martín, 2009 etc.).

Allomorphy seems to be a prime example not only of gradience and continuity of language phenomena, but also of the analogical nature of morphological production. On the one hand, allomorphic variations appertain to *the degree* of one realization or another, not to crisp, clear-cut categories. On the other hand, as humans, analogical inferences can naturally produce form variation. Based on analogy, forms can be generated in fine-grained varieties, without the need for categories (cf., Albright & Hayes, 2003).

Acknowledgments: This work was partially supported by the Ministry of Science and Environmental Protection of the Republic of Serbia (grant number: 149039D). The authors thank Tamara Jovanović for generosity in consenting large parts of data from her behavioral study, to be used here for comparisons with computer simulated outcomes. Also, the authors thank Dániel Vásárhelyi and one anonymous reviewer for their constructive criticism of an earlier version of this paper.

References

- Albright, Adam, & Hayes, Bruce (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, *90*, 119-161.
- Allen, Mark, & Badecker, William (1999). Stem homograph inhibition and stem allomorphy: Representing and processing inflected forms in a multilevel lexical system. *Journal of Memory and Language*, *41*, 105-123.
- Baayen, R. Harald (2003). Probabilistic Approaches to Morphology. In Rens Bod, Jennifer Hay, & Stefanie Jannedy (Eds.), *Probabilistic Linguistics* (pp. 229-287). Cambridge: MIT Press.

- Baayen, R. Harald, Davidson, D. J. and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412.
- Baayen, R. H., Feldman, L.B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290–313.
- Berko, J. (1958). The child's learning of English morphology. *Word*, 14, 150–177.
- Blevins, J. (2004). Inflectional classes and economy. In L. Gunkel, G. Müller & G. Zifonun (Eds.), *Explorations in Nominal Inflection* (pp. 41-85). Berlin: Mouton de Gruyter.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehard and Winston.
- Boudelaa, S., & Marslen-Wilson, W. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, 81, 65-92.
- Boudelaa, S., & Marslen-Wilson, W. D. (2004). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, 90, 106-116.
- Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.
- Daelemans, W., & Van den Bosch, A. (2005). *Memory-Based Language Processing*. Cambridge: Cambridge University Press.
- Daelemans, W., Berck, P., & Gillis, S. (1997). Data Mining as a Method for Linguistic Analysis: Dutch Diminutives. *Folia Linguistica*, 31, 57-75.
- De Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot. Edited posthumously by C. Bally, A. Sechehaye, and A. Riedlinger. Citation pagenumbers and quotes are from the English translation by Wade Baskin, New York: McGraw-Hill Book Company, 1966.
- Eddington, D., (2002a). Dissociation in Italian conjugations: A single-route account. *Brain and Language*, 81, 291-302.
- Eddington, D., (2002b). Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics*, 9, 49-75.
- Estes, W. K. (1994). *Classification and cognition, vol. 22 of Oxford Psychology Series*. New York: Oxford University Press.
- Hahn, U., & Nakisa, R. C. (2000). German inflection: Single route or dual route? *Cognitive Psychology*, 41, 313-360.

- Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, 29, 147-160.
- Harris, Z. S. (1951). *Methods in structural linguistics*. Chicago: University of Chicago Press.
- Harris, Z. S. (1957). Co-occurrence and transformation in linguistic structure. *Language*, 33, 283-340.
- Hay, J., & Baayen, R.H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, 9, 342-348.
- Hayes, B., & Cziráky-Londe, Z. (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology*, 23, 59-104.
- Ivić, P. (1990). *O jeziku nekadašnjem i sadašnjem /On past and contemporary language/*. Belgrade: Bigz-Jednistvo.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59, 434-446.
- Järvikivi, J., & Niemi, J. (2002a). Form-Based Representation in the Mental Lexicon: Priming (with) Bound Stem Allomorphs in Finnish. *Brain and Language*, 81, 412-423.
- Järvikivi, J., & Niemi, J. (2002b). Allomorphs as paradigm indices: On-line experiments with Finnish free and bound stems. *SKY Journal of Linguistics*, 15, 119-143.
- Järvikivi, J., Bertram, R., & Niemi, J. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, 21, 394-431.
- Johansen, M., & Palmeri, T. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, 45, 482-553.
- Jovanović, T. (2008). *Ispitivanje prirode alomorfije: grafo-fonoloski korelati alomorfije u srpskom jeziku i efekat frekvence sufiksa /Examining allomorphy: grapho-phonological correlates of allomorphy in Serbian and suffix frequency effect/*. Master thesis, University of Novi Sad, Serbia.
- Jovanović, T., Filipović Đurđević, D., & Milin, P. (2008). Kognitivna obrada alomorfije u srpskom jeziku /The cognitive processing of the allomorphy in Serbian/. *Psihologija*, 41, 87-101.

- Kertész, Z. (2003). Vowel harmony and the stratified lexicon of Hungarian. *The Odd Yearbook*, 7, 62-77.
- Keuleers, E. (2008). *Memory-Based learning of inflectional morphology*. PhD Dissertation, University of Antwerp, Belgium.
- Keuleers, E., & Daelemans, W. (2007). Memory-Based Learning Models of Inflectional Morphology: A Methodological Case Study. *Lingue e Linguaggio*, 6, 151-174.
- Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G., & Martens, E. (2007). Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology*, 54(4), 283-318.
- Kostić, Đ. (1999). *Frekvencijski rečnik savremenog srpskog jezika /Frequency Dictionary of Contemporary Serbian Language/*. Institute for Experimental Phonetics and Speech Pathology & Laboratory of Experimental Psychology, University of Belgrade, Serbia <<http://www.serbiancorpus.edu.rs/>>.
- Krott, A., Schreuder, R., Baayen, R. H., & Dressler, W.U. (2007). Analogical effects on linking elements in German compounds. *Language and Cognitive Processes*, 22, 25-57.
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Cybernetics and Control Theory*, 10, 707-710.
- Lieber, R. (1982). Allomorphy. *Linguistic Analysis*, 10, 27-52.
- Lyons, J. (1986). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- MacWhinney, B. (1975). Rules, rote, and analogy in morphological formations by Hungarian children. *Journal of Child Language*, 2, 65-77.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.
- 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.
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 63, 81-97.

- Mirković, J., Seidenberg, M., & Joanisse, M. (2009). Probabilistic Nature of Inflectional Structure: Insights from a Highly Inflected Language. Submitted to Cognitive Science.
- Norris, D. (2005). How do computational models help us build better theories? In A. Cutler (Ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones* (pp. 331-346). Hillsdale, N.J.: Erlbaum.
- Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 15, 39-57.
- Pléh, C. (1989). The development of sentence interpretation in Hungarian. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 158-184). New York: Cambridge University Press.
- Pléh, C., Lukács, A., & Racsmany, M. (2002). Morphological patterns in Hungarian children with Williams syndrome and the rule debates. *Brain and Language*, 86, 377-383.
- Ratcliffe, R. (1998). *The "broken" plural problem in Arabic and comparative Semitic: allomorphy and analogy in non-concatenative morphology*. Amsterdam: John Benjamins.
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46, 245-266.
- Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth Mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40, 99-121.
- Schreuder, R., & Baayen, R. H. (1995). Modeling Morphological Processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 131-154). New Jersey: Lawrence Erlbaum.
- Skousen, R. (2002). An overview of analogical modeling. In R. Skousen, D. Lonsdale, & D. B. Parkinson (Eds.), *Analogical modeling: An exemplar-based approach to language* (pp. 11-26). Amsterdam: John Benjamins.
- Smith, E., & Medin, D. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Spencer, A. (2001). Morphology. In M. Aronoff & J. Rees-Miller (Eds.), *The Handbook of Linguistics* (pp. 213-237). Oxford: Blackwell Publishers.

Zec, D. (2006). *Phonology within morphology in South Slavic: the case of OV augmentation*. Handouts, University of Nova Gorica, Slovenia.