THE EFFECT OF POLYSEMY ON PROCESSING OF SERBIAN NOUNS

Dušica Filipović Đurđević

Department of Psychology, University of Novi Sad

Aleksandar Kostić[1]

Laboratory for Experimental Psychology, University of Belgrade

It has been shown that while multiple unrelated meanings of a word (e.g. bank) increase processing latency, polysemy, that is multiple related word senses (e.g. paper) produce faster responses (Rodd, Gaskell & Marslen-Wilson, 2002; Klepousniotou, 2002). The goal of this study was to explore the effect of polysemy on word processing in Serbian. The outcomes of three lexical decision experiments have shown that polysemous words are processed faster. In addition, lemma frequency and number of related senses did not interact. Finally, a measure that combines lemma frequency and number of related senses into a single metric is proposed. This measure is information residual, initially applied on derivational morphology (Moscoso del Prado Martìn, Kostic & Baayen, 2004). In this study the information residual is a difference between the amount of information (bit) derived from lemma frequency and the entropy of a polysemic cluster. Since relative frequencies of different word senses of a given word in Serbian are currently not available, maximum entropy (log N) was used as an approximation. The outcome of this study indicates that cognitive system is sensitive not only to the entropy of derivational clusters, but polysemic clusters as well.

Key words*: polysemy, information theory, lexical decision*

---

[1]   Adresa autora: akostic@f.bg.ac.yu

Word ambiguity is one of the factors that influence processing of isolated words. Depending on the inter-relations of the word meanings two main forms of word ambiguity can be described. On the one hand, there is *homonymy*, which denotes words with unrelated meanings (for example "bank" – financial institution, and "bank" – river bank). On the other hand, there is *polysemy*, which denotes words with related senses, formed by extending the field of the original meaning of a given word (for example "paper" – material, and "paper" – scientific paper) /Lyons, 1977; Gortan-Premk, 2004/. Numerous researches demonstrated that words with multiple meanings are processed faster compared to unambiguous words (Azuma & Van Orden, 1997; Borowsky & Masson, 1996; Hino, Lupker & Pexman, 2002; Hino & Lupker, 1996). A more detailed investigation of the processing of the two forms of ambiguity revealed that polysemy decreases, while homonymy increases processing time (Rodd, Gaskell & Marslen-Wilson, 2002; Klepousniotou, 2002). Since word ambiguity has not been subjected to experimental research in Serbian language, we conducted a study on processing of Serbian polysemous words. Having in mind general similarity between polysemy, and derivation, we proposed *information residual* as a measure of word ambiguity. This measure, which will be described in more detail in the following text, has proved to be a significant predictor of processing time of words with derivational suffixes (Mosocoso del Prado Martin, Kostić & Baayen, 2004).

THE EFFECT OF POLYSEMY

Research of processing of ambiguous words demonstrated that an increase in number of meanings is followed by a decrease in processing time in a lexical decision task (Azuma & Van Orden, 1997; Borowsky & Masson, 1996; Hino, Lupker & Pexman, 2002; Hino & Lupker, 1996). However, in previous research, the difference between homonymy and polysemy was neglected. Recent research revealed that decrease in reaction time is bound to polysemy, that is to processing words with related senses (Rodd, Gaskell & Marslen-Wilson, 2002; Klepousniotou, 2002). Rodd and colleagues manipulated both number of meanings/senses, and type of ambiguity (homonymy/polysemy). In their experiment, homonymous words were processed more slowly than unambiguous words, while number of related senses (polysemy) facilitated processing (Rodd et al. 2002). The same design and the same stimuli were presented in a MEG study (Beretta, Fiorentino & Poeppel, 2005). The following of the component of magnetoencephalogram that is commonly associated with lexical activation (M350) confirmed the pattern of results observed by Rodd et al. (2002). Having in mind processing differences between two types of ambiguity, we will restrict our research to investigating the effect of polysemy.

Much attention in ambiguity processing research has been dedicated to a relation between number of meanings, and lemma frequency. However, these investigations led to conflicting results. On the one hand, in a lexical decision experiment, Jastrzembski observed a stronger ambiguity effect for the low frequency words

(Jastrzembski, 1981). On the other hand, more recent studies demonstrated that the effects of the two variables on lexical decision times were independent (Hino & Lupker, 1996; Hino, Lupker, Sears & Ogawa, 1998). However, our interest in the relation of the two variables is inspired by application of information residual as a measure for describing polysemy.

INFORMATION RESIDUAL

Early quantitative research of language revealed that words occurring more often tend to have more meanings (Zipf, 1945). A similar tendency was observed in the research of derivational morphology – the more frequent a lemma, the larger the number of the derivatives based on that word (Moscoso del Prado Martìn, Kostić & Baayen, 2003). Considering the fact that collinearity of predictors poses a problem in statistical data analysis, numerous variable control techniques are being applied. One of the solutions to this problem is achieved by finding a unique predictor based on the combination of correlated predictors. By applying this principle, Moscoso del Prado Martìn and colleagues approached the investigation of the derivational family size effect by proposing a new information theory measure, the so-called "information residual" (equation 1) /Mosocoso del Prado Martin, Kostić & Baayen, 2004/.

$$I_{residual} = I - \sum_{j=1}^{m} H_j \qquad (1)$$

Information residual (equation 1) represents a difference between the amount of information derived from lemma frequency (calculated as proportion) /equation 2/ and the sum of the entropies of morphological paradigms of the given word (equation 3)[2]. The higher the amount of information carried by a certain lemma, the longer the processing time. Whereas the higher the sum of the entropies of the words' paradigms, the shorter the processing time. Consequently, the effect of information residual would represent a *resultant* of the effect of the lemma information load, and of the effect of the sum of the entropies of it morphological paradigms.

$$I = -\log_2 p \quad (2) \qquad\qquad H_j = -\sum_{i=1}^{n} p_i \log_2 p_i \quad (3)$$

---

[2] The original research was conducted in Dutch, which is not a language of rich inflectional morphology. Therefore, only the derivational paradigms were taken into account, although the equation would predict taking into account inflectional paradigms, as well (for example, different inflected forms of a noun of a certain gender).

It should be noted that certain similarities exist between polysemy, and derivational paradigms that enable us to apply the information residual in describing polysemy. First similarity concerns the nature of derivational paradigms, and polysemy: both are based on the extension of the word's semantic field. In case of derivation, semantic variation is marked by a derivational affix, while in case of polysemy, semantic change is revealed strictly through context. The second similarity can be observed in the empirical findings, recorded in the research of the effects of the size of derivational paradigms, on the one hand, and the effects of polysemy, on the other hand. It is demonstrated that the size of the derivational paradigm (family size), that is the number of words that can be derived from a given lemma, is inversely correlated with processing time (Schreuder & Baayen, 1997). Polysemy affects processing time in a similar way (Borowsky & Masson, 1996; Hino & Lupker, 1996; Hino, Lupker & Pexman, 2002). In addition, it is shown that facilitatory effect of the number of derivatives is restricted to the derivatives that are semantically related to a given lemma, while the number of unrelated derivatives inhibits processing (Moscoso del Prado Martin, Deutsch, Frost, Schreuder, De Jong, & Baayen, 2005). Likewise, polysemy facilitates, while homonymy inhibits processing (Rodd et al., 2002; Klepousniotou, 2002).

When applied to polysemy, information residual represents a difference between the amount of information derived from lemma frequency (calculated as proportion) /equation 2/ and the sum of entropies of the clusters of related senses of a given word. Considering the lack of data on the probabilities of individual senses of a word, in this research the sum of the entropies of the sense clusters will be approximated by maximum entropy that is the logarithm of the number of senses (equation 4).

$$I_{residual} = I - \log_2 N \quad (4)$$

The amount of information based on lemma frequency (I) is positively correlated with reaction time, while, by analogy with the effect of the size of the derivational cluster, maximum entropy of the polysemic cluster should be negatively correlated with reaction time. There are two advantages to applying of the information residual in describing polysemy. On the one hand, this descriptor represents a potential solution to collinearity problem. On the other hand, if it could be demonstrated that information residual represents a cognitively relevant description of the complexity of the polysemous words, information residual would have a more global application in description of various aspects of the language.

We conducted three lexical decision experiments in order to explore the effect of number of related senses, and the possibility of applying information residual in describing polysemy. In the first experiment, we compared the groups of polysemous, and words with only one sense. Given the contradictions in the results of experiments exploring the relation between ambiguity, and lemma frequency effects, we manipulated the two in a factorial design. Finally, in the third experiment, we explored the relation between number of senses, and processing time in more detail.

EXPERIMENT 1

The main goal of this experiment was to explore the effect of polysemy that is the number of related senses on processing time, and accuracy. Participants were presented with two groups of verbs: verbs with only one sense, and verbs with as many senses as possible, given the restriction of matching the two groups for lemma frequency.

Method

Participants: Twenty-nine first-year students from The Department of Psychology, at the Faculty of Philosophy in Belgrade participated in the experiment.

Stimuli: Thirty Serbian verbs and thirty Serbian pseudoverbs were presented. Number of senses was determined based on the Rečnik Matice srpske dictionary, and frequency counts were based on Frequency Dictionary of Serbian Language (Kostić, 1965).

Design: Two factors were manipulated in the experiment: lexicality (word, pseudoword) and polysemy (unambiguous word, polysemous word). Both factors were repeated by participants, and unrepeated by stimuli. Only words were included in the analyses. Dependent variables were reaction time (in milliseconds), and percent of errors. Two groups of stimuli were matched for lemma frequency, word length in letters, and number of syllables. The description of the two groups of stimuli are presented in Table 1.

Procedure: Stimuli were presented in a visual lexical decision task. They remained on the screen until response, or time limit of 1500 ms. Prior to experiment, participants were presented with a trial session of 4 verbs, and 4 pseudoverbs.

Table 1: Average number of senses, average lemma frequency, and average word length in letters of the two groups of the stimuli presented in Experiment 1.

	Average number of senses	Average lemma frequency	Average word length in letters
Unambiguous words	1	362.1	5.7
Polysemous words	15.1	362.5	5.1

Results and discussion

Average reaction times and error percentages are presented in Figure 1. For each participant, within each cell of the experimental design, we excluded reaction times that were outside of -2/+2 standard deviation units interval. By-participant analysis of variance of reaction time revealed significant effect of polysemy: $F(1,28)=64.030$, $p<0.01$. The same effect was observed in by-participant analysis of error percentages: $F(1,28)=17.402$, $p<0.01$. However, in by-item analysis of variance none of the effects was significant.

Figure 1: Average reaction time (left plot), and average percentage of errors for the two groups of stimuli presented in Experiment 1.
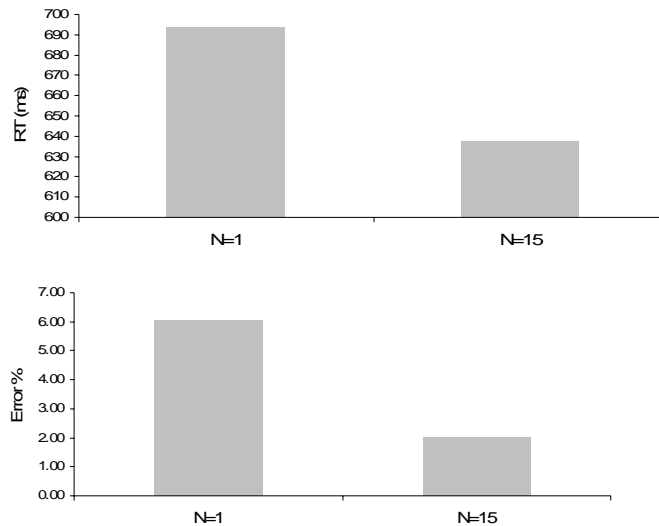
The results demonstrate that words with large number of related senses tend to be processed faster, and more accurately than words with only one sense.

EXPERIMENT 2

Experiment 2 aimed at determining the relation between lemma frequency, and number of senses. Participants were presented with four groups of Serbian verbs. Lemma frequency and number of senses were combined in a 2x2 factorial design.

Method

Participants: Twenty-seven first-year students from The Department of Psychology, at the Faculty of Philosophy in Belgrade participated in the experiment.

Stimuli: Sixty Serbian verbs and 60 Serbian pseudoverbs were presented. Number of senses was determined based on the Rečnik Matice srpske dictionary, and frequency counts were based on Frequency Dictionary of Serbian Language (Kostić, 1965).

Design: Three factors were manipulated, but only two were included in the analysis. Factor lexicality had two levels (word, pseudoword), but only words were included in the analyses. The second factor was lemma frequency (low frequency words, high frequency words), and the third factor was polysemy (unambiguous word, polysemous word). All factors were repeated by participants, and unrepeated by stimuli. Dependent variables were reaction time (in milliseconds), and percent of errors. The four groups of stimuli were matched for word length in letters. The description of the four groups of stimuli are presented in Table 2.

Procedure: The same procedure was used as in Experiment 1. Prior to experiment, participants were presented with a trial session consisting of 8 verbs, and 8 pseudoverbs.

Table 2: Average number of senses, average lemma frequency, and average word length in letters of the two groups of stimuli presented in Experiment 2.

	Low frequency		High frequency	
	Unambiguous words	Polysemous words	Unambiguous words	Polysemous words
Average number of senses	1	3.7	1	3.7
Average lemma frequency	1	1	362.1	362.3
Average word length in letters	2.9	2.9	2.7	2.9

Results and discussion

Average reaction times, and error percentages are presented in Figure 2. For each participant, within each cell of the experimental design, we excluded reaction times that were outside of -2/+2 standard deviation units interval. Analysis of variance of reaction time revealed significant main effects of lemma frequency: $F(1,26)=198.96$, $p<0.01$ (by-participant), $F(1,56)=184.87$, $p<0.01$ (by-item), and number of senses: $F(1,26)=28.384$, $p<0.01$ (by-participant), $F(1,56)=14.711$, $p<0.01$ (by-item). The interaction was not statistically significant. In error analysis, there was a significant main effect of lemma frequency: $F(1,26)=88.511$, $p<0.01$ (by-participant), $F(1,56)=19.486$, $p<0.01$ (by-item), and number of senses: $F(1,26)=17.370$, $p<0.01$ (by-participant), $F(1,56)=4.244$, $p<0.05$ (by-item), while interaction was not significant in this analysis, neither. Participants responded faster, and more accurately to high frequency words, and words with many senses.

Figure 2: Average reaction time (left plot), and average percentage of errors for the two groups of stimuli presented in Experiment 2.
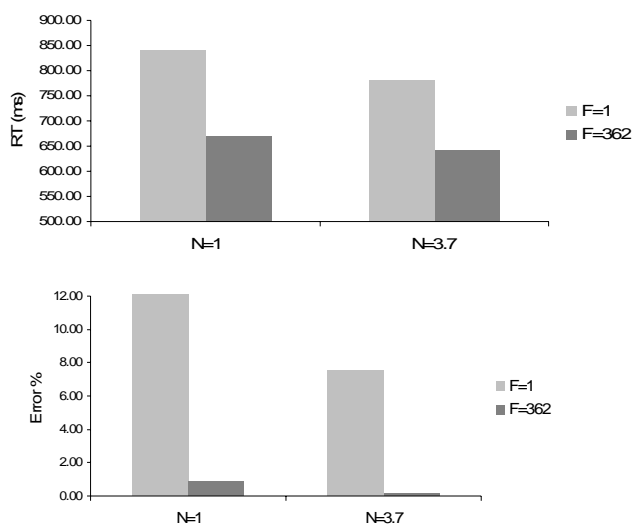
If we compared the results of Experiment 2 with the results of the first experiment, we would notice that the effect of number of senses is smaller in the second experiment. Having in mind that this difference is followed by a smaller difference in number of senses, we could pose a question on the form of functional relation between number of senses, and processing time. Therefore, in Experiment 3 we presented words with larger range in number of senses, and minimum differences in number of senses between the subsequent groups in order to explore the nature of the number of senses effect on processing time, and accuracy.

EXPERIMENT 3

Results of Experiment 1, and Experiment 2 enabled us to pose a question of the nature of relation between number of senses, and processing time. In order to interpolate, that is to describe a functional relation between the two variables, we needed a larger range of number of senses, and smaller differences between subsequent levels of the number of senses. In this experiment, reaction time, and error rates were measured on nine levels of number of senses. In addition, the findings of the previous experiment in relation to facilitatory effect of lemma frequency enabled the application of information residual as a measure of complexity of polysemous words.

Method

Participants: Twenty-four first-year students from The Department of Psychology, at the Faculty of Philosophy in Belgrade participated in the experiment.

Stimuli: Ninety Serbian nouns and ninety Serbian pseudonouns were presented. Number of senses was determined based on the Rečnik Matice srpske dictionary, and frequency counts were based on Frequency Dictionary of Serbian Language (Kostić, 1965).

Design: Two factors were manipulated in the experiment: first factor, lexicality had two levels (word, pseudoword), but only words were included in the analyses. The second factor, number of senses, had nine levels (one sense, two senses, nine senses). Both factors were repeated by participants, and unrepeated by stimuli. Lemma frequency was kept constant as much as possible, considering the general collinearity of frequency, and number of senses in language. Although there were differences in lemma frequencies across the four levels of number of senses, they showed not to be significant in analysis of variance. Dependent variables were reaction time (in milliseconds), and percent of errors. The four groups of stimuli were matched for word length in letters. The description of the four groups of stimuli are presented in Table 3.

Procedure: The same procedure was used as in Experiment 1, and Experiment 2. Prior to experiment, participants were presented with a trial session consisting of 9 nouns, and 9 pseudonouns.

Table 3: Average number of senses, average lemma frequency, and average word length in letters of the two groups of stimuli presented in Experiment 3.

Number of senses	1	2	3	4	5	6	7	8	9
Average frequency	534.3	656.6	680.1	671.5	675.1	691.8	800.6	737.2	575.5
Word length (letters)	5	5	5	5	5	5	5	5	5

Results and discussion

Average reaction times and error percentages are presented in Figure 3. For each participant, within each cell of the experimental design, we excluded reaction times that were outside of -2/+2 standard deviation units interval. By-participant analysis of variance of reaction time revealed significant main effects of number of senses: $F(8,184)=4.142$, $p<0.01$, while this effect was at the very limit of significance in by-item analysis of variance: $F(8,81)=2.010$, $p=0.05$. In error analysis, the effect of number of senses was significant only in by-participant analysis: $F(8,184)=4.058$, $p<0.01$. In addition, a linear regression analysis on nine average reaction times revealed that number of senses accounted for 49% of the processing latency variance: $F(1,7)=6.821$, $p<0.05$. Proportion of variance of error percentages was not accounted for significantly by number of senses (however, if we excluded average reaction time for words with two senses, number of senses would account for 64% of error percent variance: $F(1,7)=10.539$, $p<0.01$). Since the nine groups

were matched for lemma frequency, the effect of frequency was not significant in this analysis.

Figure 3: The relation between number of senses and reaction time (left), and number of senses and error percentages (right).
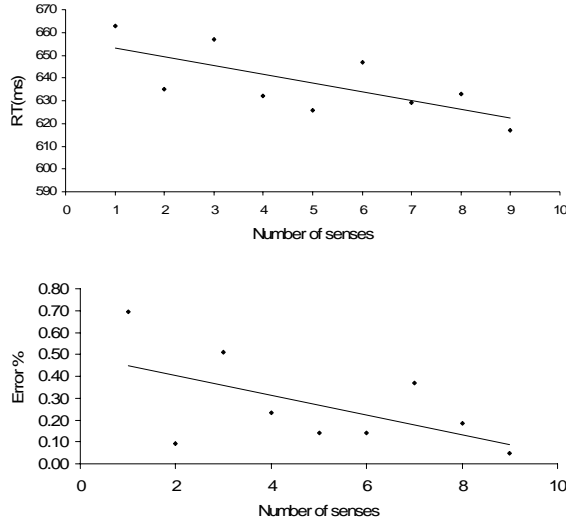
Linear regression analysis was applied to a full set of 90 points, as well. In this analysis we evaluated effects of three predictors of the reaction latencies: (log) lemma probability, number of senses, and information residual, which was calculated by subtracting (log) number of senses from (log) lemma probability (derived from lemma frequency) for each of the presented stimuli (Figure 4). Results revealed that significant proportion of reaction time variance was accounted for by (log) lemma probability: $r^2=0.108$, $F(1,88)=10.604$, $p<0.01$, as well as (log) number of senses: $r^2=0.084$, , $F(1,88)=8.054$, $p<0.01$. In multiple regression analysis (log) lemma probability, and (log) number of senses taken together, accounted for 22% of variance: $r^2=0.223$, $F(2,87)=12.517$, $p<0.01$. The same proportion of variance was accounted for by information residual, which was calculated by combining the two predictors into a single measure: $r^2=0.221$, $F(1,88)=24.933$, $p<0.01$.

Figure 4: Relation between reaction time and (log) lemma frequency (upper left), number of senses (upper right), and information residual (lower).
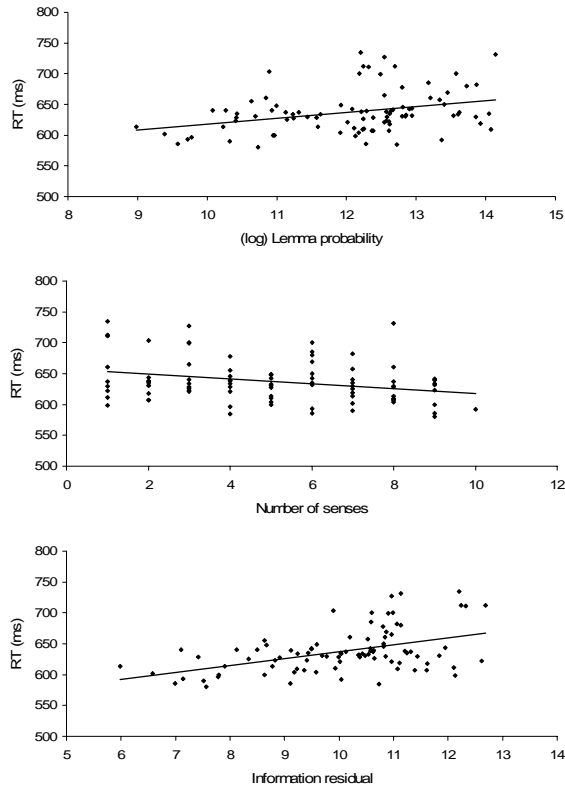
   The results demonstrated that an increase in number of senses was followed by a decrease in reaction time. Based on this, it can be deduced that processing time would be inversely proportional to a maximum entropy of the sense probability distribution, as well[3]. At the same time, processing time increased with an increase in the logarithm of lemma probability that is amount of information derived from lemma frequency. Based on the directions of these effects, we could deduce that the two predictors could be combined in a single measure of information residual, which was confirmed in the results of multiple regression analysis.

---

[3]   Maximum entropy, as the logarhitm of the number of senses is monotonically increasing function of the number of senses.

GENERAL DISCUSSION

We conducted three experiments to explore the effect of polysemy in processing of Serbian language. In addition to polysemy (number of senses), we looked at it's relation to lemma frequency, and suggested a unique measure which would combine the two.

In the first experiment, unambiguous, and polysemous verbs were compared for processing latencies, and errors. The two groups were matched for word length in letters, and lemma frequency, and were selected to represent the two extremes on the number of senses continuum. The results revealed the processing advantage of the polysemous verbs.

In the second experiment, groups of verbs were selected to have either one, or many senses, and either low or high lemma frequency. Fifteen polysemous nouns were selected to fill each cell of the factorial design, in such a way that the number of senses was matched across each level of lemma frequency, and lemma frequency was matched across each level of number of senses. The four groups of words were matched for length in letters, and number of syllables. The results revealed a processing advantage for the verbs with many senses, and verbs of high lemma frequency. Although the effect of number of senses was to a certain extent stronger in case of low frequency verbs, the interaction of the two factors was not statistically significant. The observed results are in accordance with recent studies conducted in English (cf. Hino & Lupker, 1996).

Having in mind that the reduced difference in average number of senses between the unambiguous words, and polysemous words presented in two experiments was followed by the reduced difference in processing time we opened a question of the more detailed description of the number of senses effect. Therefore, in Experiment 3, we presented nine groups of nouns that were matched for lemma frequency, and differed in number of senses. Results of linear regression analysis revealed that number of senses accounted for significant proportion of processing time variance. However, significant deviations from the predicted values suggest that number of senses based on Rečnik Matice srpske dictionary was not a reliable estimate. Thus, the future research would aim at finding alternative ways in estimating the number of senses.

One of the main problems in studies of polysemy effect is a high correlation between number of senses, and lemma frequency. A similar problem was encountered in the study of the effect of derivational family size (cf. Moscoso del Prado Martin, Kostić & Baayen, 2003). Moscoso del Prado Martin, and colleagues proposed a solution to collinearity problem; they suggested that derivational family size, and lemma frequency should be combined in a single measure of information residual. This measure is calculated as a difference between the amount of information based on lemma frequency, and the sum of the entropies of the derivational

paradigms of a given word. The effect of information residual presented a resultant of the inhibitory effect of lemma's information load, and the facilitatory effect of the sum of the entropies of the word's derivational paradigms.

There are certain similarities between polysemy, and morphological family size. On the one hand, both number of senses, and morphological family size are negatively correlated with processing time (Hino & Lupker, 1996; Schreuder & Baayen, 1997). On the other hand, an inhibitory effect of the unrelated meanings, and a facilitatory effect of the related meanings is observed, both in word ambiguity processing, and in derivational morphology. (Rodd et al., 2002; Moscoso del Prado Martìn et al., 2005). Having in mind the observed similarities, we suggested to solve the collinearity problem by combining lemma frequency, and number of senses in a single measure (information residual). Lacking the data for probabilities of the individual senses, entropy of the sense probability distribution was approximated by maximum entropy that is by logarithm of the number of senses. The results of Experiment 3 demonstrated that information residual accounted for the same proportion of reaction time variance as did lemma frequency, and number of senses taken together. The advantage of information residual is expected in the studies dealing with material of high collinearity between lemma frequency, and number of senses. Based on the results of our study, we concluded that the measure of information residual, which was initially proposed as a measure of morphological complexity of a word, could also be successfully applied in describing cognitive complexity of polysemous words. Based on the mentioned similarities between derivational morphology, and ambiguity, we would predict a successful application of the appropriate modification of this measure to homonymy, or words with unrelated meanings, as well.

The main weaknesses of the current study are related to the way of estimation of the number of senses, on the one hand, and of the entropy of sense probability distribution, on the other hand. We based our estimations of number of senses on Rečnik Matice srpske unabridged dictionary. Some investigations have demonstrated that number of dictionary senses does not represent cognitively relevant measure of ambiguity (Gernsbacher, 1984; Lin & Ahrens, 2005). As the most common argument against relying on dictionary senses, it is stated that dictionaries often encompass senses that are unfamiliar to the majority of the speakers, and miss listing senses that the speakers are highly familiar with. However, there are studies in which ambiguity effect was demonstrated based on dictionary senses (Jastrzembski, 1981; Rodd, Gaskell & Marslen-Wilson, 2005; Beretta, Fiorentino & Poeppel, 2005). The results of our research demonstrated that dictionaries could provide a rough approximation of the number of senses. This approximation could be enough for comparing the groups of unambiguous and highly polysemous words in a factorial design. The limits of dictionary-based approximations would manifest if number of senses would be approached as a continuous variable, in a correlation design.

Another weakness of the current research is related to the entropy estimation. Lacking the data on probabilities of individual senses, we approximated entropy of sense probability distribution by a logarithm of the number of senses. Consequently, we treated all of the polysemous words as words with maximum entropy of sense probability distribution, that is, as words with a given number of equally frequent

senses. By doing so, we lost information on the actual sense probabilities. In spite of the mentioned weaknesses, the findings of the current research indicated that the application of information residual in describing of polysemy was sound. In addition to solving of the collinearity problem, information residual could be applied to various language phenomena, which opens a way towards a better understanding of possible cognitive mechanisms involved in processing of morphologically, and semantically complex words.

REFERENCES

Azuma, T., & Van Orden, G. C. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Languag*e, *36*, 484–504.

Beretta, A., Fiorentino, R. & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research*, *24*, 57-65.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognitio*n, *22*,63–85.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: Genera*l, *11*3, 256–281.

Gortan-Premk, D. (2004). *Polisemija i organizacija leksičkog sistema u srpskome jeziku*. Beograd, Zavod za udžbenike i nastavna sredstva.

Hino, Y., & Lupker, S. J. (1996). The effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performanc*e, *22*, 1331–1356.

Hino, Y., Lupker, S. J., Sears, C. R. & Ogawa, T. (1996). The effects of polysemy for Japanese katakana words. *Reading and Writing: An Interdisciplinary Journal*, *10*, 395–424.

Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and Synonymy Effects in Lexical Decision, Naming, and Semantic Categorization Tasks: Interactions Between Orthography, Phonology, and Semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 686–713.

Jastrzembski, J. E. (1981). Multiple meanings, number of related mean-ings, frequency of occurrence, and the lexicon. *Cognitive Psycholog*y, *13*, 278–305.

Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the time-course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performanc*e, *14*, 601–609.

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Languag*e, *45*, 259–282.

Klepousniotou, E. (2002). The processing of lexical ambiguity: homonymy and polysemy in the mental lexicon. *Brain & Language*. *81*(1-3), 205-23.

Kostić, Đ. (1999). *Frekvencijski rečnik savremenog srpskog jezika*. Tom I – VII. Beograd, Institut za eksperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju Filozofskog fakulteta u Beogradu.

Lin, C. C., & Ahrens, K. (2005). How Many Meanings Does A Word Have? Meaning Estimation in Chinese and English. In J. W. Minett & W. S-Y. Wang (Eds.) *Language Acquisition, Change and Emergence—Essays in Evolutionary Linguistics*. Hong Kong City, University of Hong Kong Press.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge, UK: Cambridge Univ. Press.

Mills, M. L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition*, *17*,141-147.

Moscoso del Prado Martìn, F., Kostić, A. & Baayen, H. (2003). Putting the bits together: An information-theoretical perspective on morphological processing. *Paradigmatic Structures in Morphological Processing: Computational and Cross-Linguistic Experimental Studies*. Wageningen, Ponsen & Looijen bv.

Moscoso del Prado Martìn, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H. & Baayen, H. (2005). Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language*, *53*, 496–512.

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.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behaviour*, *9*, 487–494.

Schreuder, R. & Baayen, R.H. (1997). How complex simplex words can be. *Journal of Memory and Language*, *37*, 118–139.

Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, *33*, 251-256.

APPENDIX

Stimuli presented in three experiments

experiment 1						
stimuli	length in letters	lemma frequency	number of senses	stimuli	length in letters	lemma frequency	number of senses
padati	6	961	1	naći	4	1031	23
davati	6	719	1	poći	4	733	15
stići	5	655	1	teći	4	647	14
dizati	6	536	1	držati	6	446	16
dići	4	372	1	gaziti	6	395	11
šumeti	6	354	1	preći	5	339	23
istaći	6	279	1	bacati	6	275	17
mahati	6	240	1	izaći	5	244	13
rađati	6	240	1	tući	4	241	24
sjati	5	225	1	vezati	6	219	13
nicati	6	208	1	kucati	6	219	9
ličiti	6	200	1	jesti	5	201	9
žudeti	6	189	1	sinuti	6	186	8
čuditi	6	141	1	seći	4	143	20
cvasti	6	113	1	točiti	6	119	11

experiment 2						
stimuli	length in letters	lemma frequency	number of senses	stimuli	length in letters	lemma frequency	number of senses
čađiti	6	1	1	natući	6	1	9
dopeći	6	1	1	bubati	6	1	6
izesti	6	1	1	šenuti	6	1	4
kihati	6	1	1	čepiti	6	1	4
krkati	6	1	1	baliti	6	1	4
mititi	6	1	1	usuti	5	1	4
mršati	6	1	1	kajati	6	1	3
potaći	6	1	1	mrljati	7	1	3
rašiti	6	1	1	uzneti	6	1	3
rudeti	6	1	1	vrveti	6	1	3
smleti	6	1	1	stući	5	1	3
veriti	6	1	1	ušiti	5	1	3
zadići	6	1	1	uleći	5	1	3
zaleći	6	1	1	šiljiti	7	1	2
zataći	6	1	1	odliti	6	1	2
padati	6	961	1	menjati	7	429	9
davati	6	719	1	roditi	6	440	6
stići	5	655	1	početi	6	854	4
dizati	6	536	1	podići	6	292	4
dići	4	372	1	umreti	6	553	4
šumeti	6	354	1	kriti	5	398	4
istaći	6	279	1	uspeti	6	237	3
mahati	6	240	1	gubiti	6	271	3
rađati	6	240	1	disati	6	389	3
sjati	5	225	1	šetati	6	164	3
nicati	6	208	1	vršiti	6	410	3
ličiti	6	200	1	sanjati	7	626	3

žudeti	6	189	1	kupati	6	113	3
čuditi	6	141	1	mrzeti	6	134	2
cvasti	6	113	1	kleti	5	124	2

Experiment 3			
stimuli	length in letters	lemma frequency	number of senses
šapat	5	302	1
mržnja	5	320	1
cesta	5	391	1
čežnja	5	414	1
granje	5	424	1
dečak	5	448	1
april	5	453	1
osmeh	5	720	1
cveće	5	786	1
lišće	5	1085	1
ekipa	5	254	2
bašta	5	272	2
radio	5	317	2
pesak	5	320	2
uspeh	5	374	2
uslov	5	422	2
dinar	5	894	2
odbor	5	1052	2
senka	5	1209	2
vlada	5	1452	2
suton	5	334	3
puška	5	335	3
seljak	5	335	3
sprat	5	348	3
minut	5	414	3
vidik	5	430	3
magla	5	631	3
miris	5	657	3
trava	5	1470	3
ulica	5	1847	3
zanos	5	278	4
obzir	5	280	4
pažnja	5	295	4
svest	5	314	4
potok	5	328	4
zvono	5	375	4
pomoć	5	484	4
grudi	5	831	4
grana	5	1256	4
vetar	5	2274	4
komad	5	262	5
lopta	5	270	5
igrač	5	281	5
priča	5	409	5
savez	5	517	5

REZIME

UTICAJ POLISEMIJE NA OBRADU IZOLOVANIH REČI SRPSKOG JEZIKA

Dušica Filipović Đurđević

Odsek za psihologiju, Filozofski fakultet u Novom Sadu i

Aleksandar Kostić

Laboratorija za eksperimentalnu psihologiju, Univerzitet u Beogradu

Istraživanja izvedena na engleskom jeziku pokazala su da postojanje nepoveza-nih značenja reči (homonimija, npr. "osa" – insekt i "osa" – koordinata) usporava obradu, dok postojanje povezanih značenja reči (polisemija, npr. "glava" – deo tela i "glava" – poglavlje) ubrzava obradu (Rodd, Gaskell & Marslen-Wilson, 2002; Klepousniotou, 2002). Cilj ovog istraživanja je ispitivanje uticaja polisemije na obradu reči srpskog jezika. Rezultati tri eksperimenta sa zadatkom leksičke odluke pokazali su da se polisemične reči obrađuju brže. Pored toga, pokazano je da ne postoji inter-akcija između frekvencije leme i broja značenja. Konačno, predložena je nova mera višeznačnosti, zasnovana na ideji primene "informacionog reziduala" kao mere morfološke kompleksnosti reči (Moscoso del Prado Martin, Kostic & Baayen, 2004). U našem istraživanju informacioni rezidual predstavlja razliku između količine in-formacije izvedene iz frekvencije leme (log p) i entropije distribucije značenja reči. S obzirom na to da u ovom trenutku ne raspolažemo verovatnoćama značenja reči srpskog jezika, entropija je aproksimirana maksimalnom entropijom (log N). Kombi-novanje frekvencije leme i broja značenja u jedinstvenu probabilističku meru rešava problem kolinearnosti dve varijable. Pored toga, mogućnost primene informacionog reziduala na slučaj polisemije ukazuje na mogućnost da je kognitivni sistem osetljiv na entropiju distribucije verovatnoća značenja reči.

Ključne reči*: polisemija, informaciona teorija, leksička odluka*