

VECTOR BASED SEMANTIC ANALYSIS REVEALS ABSENCE OF COMPETITION AMONG RELATED SENSES¹

Dušica Filipović Đurđević²

Department of Psychology, University of Novi Sad, Serbia

Đorđe Đurđević

Faculty of Electrical Engineering, University of Belgrade, Serbia

Aleksandar Kostić

Laboratory of Experimental Psychology, University of Belgrade, Serbia

Previous research demonstrated that processing time was facilitated by number of related word senses (polysemy) and inhibited by number of unrelated word meanings (homonymy). The starting point of this research were the findings described by Moscoso del Prado Martín and colleagues, who offered a unique account of processing of two forms of lexical ambiguity. By applying the techniques they proposed, for the set of strictly polysemous Serbian nouns we calculated ambiguity measures they introduced. Based on the covariance matrix of the context vectors, we derived entropy of equivalent Gaussian distribution, and based on the context vectors probability density function, we derived differential entropy. Negentropy was calculated as the difference between the two. Based on interpretation that entropy of equivalent Gaussian mirrors sense cooperation, or polysemy, while negentropy mirrors meaning competition, or homonymy, we predicted that in the set of strictly polysemous nouns, negentropy effect would disappear. In accordance with our predictions, entropy of equivalent Gaussian distribution accounted for significant proportion of processing latencies variance. Negentropy did not affect reaction time. This finding is in accordance with the hypothesis that entropy of equivalent Gaussian distribution, as a measure of general width of activation in semantic space,

¹ This research was supported by the Ministry of Science and Environmental Protection of the Republic of Serbia (grant number: 149039D).

² ✉: dmfilipo@f.bg.ac.rs

reflects polysemy, that is, the existence of related senses. Therefore, polysemy advantage could be the result of the wide-spread activation in semantic space and reduced competition among overlapping Gaussians.

Key words: *polysemy, semantic space, context vectors, differential entropy, equivalent Gaussian entropy, negentropy*

Processing effects of lexical ambiguity were demonstrated in large number of empirical studies (Borowsky & Masson, 1996; Hino & Lupker, 1996; Hino, Lupker & Pexman, 2002; Millis & Button, 1989). However, depending on the nature of lexical ambiguity, an increase in number of meanings/senses could be followed by an increase, or a decrease in processing time. Regarding ambiguous words of unique spelling/sound, the crucial difference is the one between homonymy and polysemy. Homonymous words, that is, words with several unrelated meanings (e.g. *bank* – river bank, and *bank* - institution) take longer to process than unambiguous words, while polysemous words, that is, words with several related senses (e.g. *paper* – scientific paper, and *paper* - material) are processed faster than unambiguous words (Beretta, Fiorentino & Poeppel, 2005; Filipović Đurđević and Kostić, 2008; Klepousniotou, 2002; Pyllkanen, Llinas & Murphy, 2006; Rodd, Gaskell i Marslen-Wilson, 2001; 2002).

The approach that offered unique account of both homonymy and polysemy effects, without relying on prior familiarity with number of meanings/senses was proposed by Moscoso del Prado Martín and colleagues (submitted). The starting point of this approach was a multinomial distribution of context vectors of a given ambiguous word and probability density function that defines it. In the first step, by building second-order co-occurrence vectors (Shütze, 1998) for each ambiguous word, they obtained a distribution in high-dimensional space of its variation in meaning. In the next step, probability density function of the given distribution was estimated. Moscoso del Prado Martín and colleagues proposed to estimate probability density function by applying the so called, infinite mixture models. An infinite mixture model is based on the assumption that a hyper-dimensional distribution is under lied by an unknown, very large, but finite number of multinomial Gaussian distributions (Neal, 1991; 1998).

The authors have found theoretical background for their approach in the language processing theory proposed by Pulvermüller (Pulvermüller, 2001). According to this theory, processing of certain elements of language corresponds to activation of neural assemblies that are distributed over various brain regions. The neurons within an assembly are interconnected by facilitatory connections, which, in accordance with the Hebbian learning principle, arise as a consequence of the frequent shared activation (Hebb, 1949). The existence of such connections implies that activation of one of the neurons will facilitate the activation of all of the neurons within a given assembly. Unlike neurons within an assembly that are bound with facilitatory connections the

neurons in distinct assemblies are rarely simultaneously activated, which results in mutual inhibition. Mechanism of inhibition prevents simultaneous activation of more than one neuron assembly. The degree of similarity among the elements of language can be depicted by the number of shared neurons in certain assemblies, that is, the degree of overlap among the assemblies. Moscoso del Prado Martín and colleagues consider the activation of certain assemblies to reflect the activation of certain meanings during the processing of ambiguous words. In this view, processing time would reflect the time necessary for resolving multiple assemblies' activation. In case of homonymous words, there would be competition among assemblies, which would lead to longer processing time. On the other hand, in case of polysemous words, due to relatedness of senses, that is large number of shared neurons among overlapping assemblies, the competition should be reduced.

The crucial link between theory of neural assemblies and a mixture of multinomial Gaussian distributions would be the hypothesis of Moscoso del Prado Martín and colleagues that the activation of a certain neural assembly would correspond to one Gaussian distribution in a multidimensional semantic space. This hypothesis is based on the frequent finding that neurons have Gaussian-like receptive fields, as explained in detail by Moscoso del Prado Martín et al. (submitted). Activation of large number of assemblies, would lead to large number of Gaussians. A mixture of these Gaussians would constitute a complex probability density function. Processing time would be a function of the level of uncertainty that is the differential entropy of the given distribution (Shannon, 1948).

Considering the difficulties of analytic derivation of differential entropy for the mixture of Gaussians, differential entropy could be numerically estimated using Monte Carlo integration (cf. Moscoso del Prado Martín et al., submitted). Monte Carlo integration is most successful when applied to the large number of dimensions (McKay, 2003). In this method, differential entropy $h(p)$ of probability density function $p(\mathbf{x})$ was approximated as a negative value of the sums of log probabilities of each of the points from the sample of points that are distributed according to the probability density function $p(\mathbf{x})$, in an n -dimensional space (equation 1).

$$h(p) \approx -\frac{1}{N} \sum_{i=1}^N \log_2 p(\mathbf{u}_i) \quad (1)$$

However, value of the differential entropy of the mixture of Gaussians is affected by two aspects of such distribution. On the one hand, differential entropy is directly proportional to the general variability of the points in the hyper dimensional space, that is, general width of the space the points span. The larger the width of the space that the word spans, the greater is the probability of its activation. On the other hand, value of the differential entropy is affected by the number of Gaussians that underlie a given probability density function, relative probabilities of each of the Gaussians, and the degree of their overlap. Unlike general variability, which facilitates processing, presence of each of the three aspects inhibits processing. The larger the number

of Gaussians, the more equal their relative probabilities, and the less the degree of overlap, the larger is the extent of competition among them (cf. Moscoso del Prado Martín et al., submitted). Therefore, it is necessary to separate the elements which facilitate from the elements which inhibit processing. Moscoso del Prado Martín and colleagues have proposed that the general variability, which facilitates processing could be expressed as the entropy of the equivalent Gaussian distribution, that is, Gaussian distribution with the identical mean, and covariance. The value of the Equivalent Gaussian Entropy (EGE) can be estimated based on the determinant of the covariance matrix of the given set of points (equation 2).

$$h(p_N) = \frac{n}{2} \ln \Gamma_2(2\pi e) + \frac{1}{2} \ln |\mathbf{K}| \quad (2)$$

Authors have proposed negentropy as a measure which reflects those aspects of the multinomial distribution that inhibit processing. Negentropy is Information-Theory measure that reflects the level of order in the system, that is, the distance of the system from the normal distribution (cf. Moscoso del Prado Martín et al.). The value of negentropy is obtained as the difference between the entropy of the equivalent Gaussian distribution, and the differential entropy of the given multinomial distribution (equation 3).

$$J(p) = h(p_N) - h(p) \quad (3)$$

This way EGE, that is, the general width of activation and a source of facilitation in processing were separated from the negentropy, which encapsulates sources of inhibitory effect on processing time (the number of Gaussians underlying the distribution, their relative probabilities, and the level of their overlap). The authors evaluated their prediction on the set of stimuli presented in a research of Rodd et al., which demonstrated the difference between polysemy, and homonymy processing (Rodd et al., 2002). In accordance with the results of Rodd and her colleagues, as well as in accordance with their own predictions, Moscoso del Prado Martín and colleagues demonstrated that entropy of the equivalent Gaussian distribution (EGE), that is, the general width of activation was negatively correlated with reaction times, while negentropy was positively correlated with reaction times. The authors consider facilitatory effect of EGE to reflect polysemy, that is, the existence of related word senses. At the same time, they stated that this effect could be considered the same as the effect of contextual distinctiveness, proposed by McDonald (2002). On the other hand, the effect of negentropy would depict the effect of homonymy, that is, the existence of unrelated word meanings.

The method proposed by Moscoso del Prado Martín and colleagues was originally performed on the set of both polysemous, and homonymous words. The aim of the present study was to apply this method to strictly polysemous Serbian words, as an additional validation of their hypotheses. On the one hand, this would provide further support by extending the original findings to a new language. On the other hand,

it would enable us to test specific aspect of the original hypotheses, by deriving a prediction related to polysemous words: if the existence of unrelated meanings, that is, a large number of mutually distant Gaussian distributions reflected homonymy, we would expect the absence, or at least a reduction of the effect of negentropy on processing time of polysemous words. This prediction was tested using the technique which is identical to the one originally applied by Moscoso del Prado Martín and colleagues.

METHOD

Participants, stimuli and procedure: The data analyzed in this study originated from an earlier study of Filipović Đurđević and Kostić (2006). In this study, 150 strictly polysemous Serbian nouns were presented to 54 participants in a visual lexical decision task.

Design: We built second-order co-occurrence vectors for the chosen set of polysemous nouns (Schütze, 1998). Context words were 1000 most frequent words of Serbian language (based on Frequency Dictionary of Contemporary Serbian Language (Kostić, 1999). Vector representations were built using electronic text data-base of journal articles of Ebart Media Documentation (www.arhiv.co.yu), which contains 70 millions of words. For each of 150 polysemous words, we formed a matrix with columns representing context words, and rows representing individual occurrences of a given polysemous word. After that, we selected 130 words with more than 500 occurrences in the text data-base. We randomly selected 500 occurrences of each polysemous word. This way, we ended up with 65 000 contextual vectors in total. In order to reduce the number of dimensions, and to deal with sparse data problem, the context vectors were subjected to a PCA (after centering to zero, and scaling their components to unit variance). In order to speed the calculation of the PCA matrix of loadings (also known as rotation matrix /Baayen, 2008/), we randomly selected a subset of 50 occurrences of each of the target words, and the matrix of loadings was computed on this smaller sample of 6 500 context vectors. This way, number of dimensions was reduced from 1000 to four principal components, which accounted for more than 90% of the variance (with no additional component accounting for more than 5% of variance). The resulting matrix of loadings was applied to the full set of vectors, for each of 130 words, and for each word four principal components were selected. After that, all of the vectors were transformed to unit length, by dividing each of the vector components by the length of the vector. This transformation was necessary in order to eliminate the effect of the differences among word frequencies. Finally, by applying the software for flexible Bayesian modeling, for each polysemous word, we estimated multinomial probability density function, that is, number of the multinomial Gaussian distributions, and their parameters (relative probability, mean, and variance for each of the four dimensions) /FBM; Neal, 2004/. Applying the same software, based on

the obtained parameters, we estimated probability of each of the 500 points, for each of the polysemous words. For these points, differential entropy of multidimensional distribution was approximated numerically, using Monte Carlo Integration (equation 1). For the same set of points, we calculated covariance matrix, based on which we estimated differential entropy of the corresponding Gaussian distribution (equation 2). Negentropy was calculated as a difference between differential entropy of the corresponding Gaussian distribution, and the differential entropy of a mixture of Gaussian distribution (equation 3). These measures were predictors in regression analysis, along with word length in letters, log lemma frequency, word familiarity and entropy of discrete sense probability distribution (Filipović Đurđević and Kostić, 2006). Dependent variables were reaction time (in milliseconds), and error counts.

RESULTS

Prior to analysis, items that elicited above 20% errors were excluded from the data set. The same criterion was applied to participants. Finally, all data points exceeding the range of ± 2.5 units of standard deviation in a distribution of reaction times were excluded as well. Reaction times were logarithmically transformed to correct for the asymmetry in the distribution.

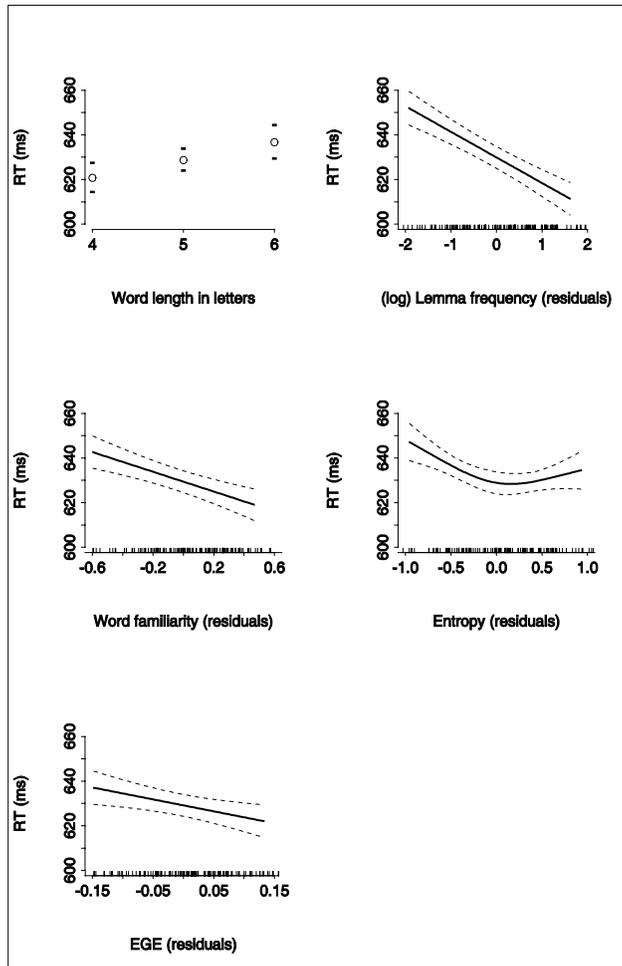
Given the co-linearity between word length in letters and lemma frequency ($r = -0.16$, $p = 0.07$), between log lemma frequency and word familiarity ($r = 0.39$, $p < 0.01$), between log lemma frequency and entropy of discrete sense probability distribution ($r = 0.23$, $p < 0.05$), as well as between entropy of discrete sense probability distribution and EGE ($r = -0.22$, $p < 0.05$), all of the predictors were decorrelated prior to analysis. The first step in this analysis was to partial out the variance of log lemma frequency that could not be accounted for by word length in letters in a linear model, to partial out the variance of word familiarity that could not be accounted for by log lemma frequency in a linear model, to partial out the variance of discrete entropy that could not be accounted for by log lemma frequency in a linear model, and to partial out the variance of EGE that could not be accounted for by discrete entropy in a linear model.

Linear mixed effect analysis of log reaction times, with participants and items as random effects, and word length in letters, log lemma frequency residuals, word familiarity residuals, entropy of discrete sense probability distribution residuals, Entropy of Equivalent Gaussian distribution (EGE), and negentropy of probability density function of context vectors as fixed effects, revealed significant linear effects for word length in letters ($\beta = 0.014$, $t(1,6573) = 3.07$, $p < 0.01$), log lemma frequency residuals ($\beta = -0.022$, $t(1,6573) = -7.89$, $p < 0.001$), word familiarity residuals ($\beta = -0.043$, $t(1,6573) = -5.13$, $p < 0.001$), discrete entropy residuals ($\beta = -0.014$, $t(1,6573) = -2.62$, $p < 0.01$), nonlinear effect of discrete entropy residuals ($\beta = 0.027$, $t(1,6573) = 3.78$, $p < 0.001$), and a linear effect of Entropy of Equivalent Gaussian distribution – EGE

residuals ($\beta=-0.084$, $t(1,6573)=-2.46$, $p=0.01$). In addition, there was an interaction between discrete entropy residuals and familiarity residuals ($\beta=0.051$, $t(1,6573)=2.70$, $p<0.01$). Neither the main effect of negentropy, nor the interaction between negentropy and EGE were significant. There were no interactions among fixed and random effects. Figure 1 depicts partial effects of all predictors from the model.

A logistic regression to the number of correct and incorrect responses revealed significant effects of log lemma frequency residuals ($\chi^2=44.35$, $p<0.001$), word familiarity residuals ($\chi^2=25.24$, $p<0.001$), and entropy of discrete probability distribution ($\chi^2=12.38$, $p<0.001$). Entropy of Equivalent Gaussian distribution, nor negentropy of probability density function of context vectors had no effect on error probability.

Figure 1: Partial effects of word length residuals, (log) lemma frequency residuals, word familiarity residuals, residuals of entropy of discrete sense probability distribution and EGE residuals.



DISCUSSION

Our research directly relied on the findings described by Moscoso del Prado Martín and colleagues, which demonstrated that vector based semantic measures could not only serve as a way of quantification of semantic variables, as suggested earlier (Landauer and Dumais, 1997; Lund and Burgees, 1997; McDonald, 2000; Schütze, 1998), but could also describe the differences between homonymy and polysemy (Moscoso del Prado Martín, 2006; Moscoso del Prado Martín et al., submitted). In the referred paper, by analyzing a set of both homonymous and polysemous English words, the authors came to a conclusion that facilitatory effect of polysemy was a consequence of general width of distribution, that is, a general span of a given word's distribution of context vectors in the semantic space. General width of distribution was expressed by Entropy of Equivalent Gaussian distribution derived from the covariance matrix of the given set of points in the hyper dimensional space. On the other hand, inhibitory effect of homonymy was interpreted as a consequence of competition among distant and unequally distributed Gaussians that represented unrelated meanings. These Gaussians were estimated as components of a complex, multinomial probability density function of the given set of points. The sources of inhibitory effects were represented as negentropy of a mixture of Gaussians. Moscoso del Prado Martín and colleagues, in light of Pulvermüller's theory (Pulvermüller, 2003), consider that faster processing of polysemous words could be accounted for by higher initial level of activation of words that appear in high number of various contexts, that is, occupy larger proportion of semantic space. On the other hand, negentropy depicts the level of competition. Moscoso del Prado Martín and colleagues state that the level of competition is related to homonymy, that is, unrelated word meanings, which occupy distant parts of semantic space, and which do not overlap. High degree of competition leads to longer processing time of homonymous words. By this account, faster processing of polysemous words could be a consequence of the wide distribution in semantic space, on the one hand, and the absence of competition, on the other hand.

Based on their conclusions, we predicted that in the absence of unrelated meanings there would be no effect of negentropy, that is, that there would only be the facilitatory effect of general width of distribution present. In order to test this prediction, we applied their techniques to the set of strictly polysemous Serbian nouns. Polysemous words being words with related senses, we predicted only the effect of general width of distribution, or EGE.

As predicted, the results revealed a facilitatory effect of Entropy of Equivalent Gaussian distribution, which indicated that the words appearing in the larger number of contexts were processed faster. However, given the nature of our stimuli, this conclusion should be limited to the contexts that are mutually related. There was no main effect of negentropy in the given set of polysemous words, which indicated the absence of competition among unrelated meanings.

The absence of negentropy effect in case of processing strictly polysemous words, which was obtained in our study, could be of importance for understanding the way two sorts of ambiguity are represented. Moscoso del Prado Martín and colleagues stated three possible sources of competition: a) the number of Gaussians that underlie the probability density function, b) unequal levels of their relative probabilities, and c) the degree of their overlap (Moscoso del Prado Martín et al., submitted). Based on the absence of negentropy effect in processing of polysemous words, it is not possible to rule out separately the effects of each of the three inhibitory factors. The absence of competition among the neural assemblies could suggest that polysemous words are represented by single neural assembly. Another possibility would be that the senses of the polysemous words are represented by distinct neural assemblies, as in the case of the meanings of homonymous words. In this case, faster processing of polysemous words could be accounted for by extremely high degree of overlap among the neural assemblies representing word senses, that is, a large number of shared neurons in these assemblies, leading to a high level of mutual facilitation.

These findings could also be interpreted in light of the model proposed by Rodd and colleagues (Rodd et al., 2004). According to this model, homonymous words form narrow, deep attractor basins which occupy distant parts in semantic space, while polysemous words form wide, shallow attractor basins in one part of the semantic space. Consequently, the time it takes for the network to settle in one pattern of activation is shorter for polysemous words. However, Moscoso del Prado Martín and colleagues (submitted) state that, in addition to the distance among the meanings in semantic space, competition can be under influence of the balance of probabilities of the meanings. According to them, there is a high level of competition in case of a word with equally frequent meanings, due to long time needed for resolving the activation (in case of unequal probabilities, the dominant meaning inhibits subordinate meanings in short time).

It should be noted that entropy derived from the distribution of context vector probabilities did not eliminate the effect of entropy of discrete sense probability distribution, as described by Filipović Đurđević and Kostić (2006). Therefore, an open question remains of the relationship between traditional measures of ambiguity, and the measures derived from the context vector probability distribution. McDonald stated that the ambiguity can be reduced to the width of activation, since the results of his studies demonstrated that the effect of number of meanings disappeared when the words were previously matched for the level of contextual distinctiveness (McDonald, 2002). Our results did not demonstrate this. A possible reason for the deviation could be assigned to a large number of approximations that occurred during the estimation of the probability density function, as well as to the choice of the context words. Regardless, one should be particularly careful when considering the effects of the measures derived from the distribution of discrete probabilities obtained by the participants, on the one hand, and the effects of measures derived from the distribution of the context vectors, on the other hand. Therefore, the future research should aim at the detailed study of the relation among the measures based on the quantitative analysis

of language corpora, on the one hand, and the measures based on the semantic intuition of the native speakers, on the other hand. The full understanding of the relation among these measures would enable us to formulate the unique principle in accounting for the influence of different forms of word ambiguity, such as polysemy, homonymy, homography, homophony and so forth.

REFERENCES

- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- 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 Cognition*, 22, 63–85.
- Filipović Đurđević, D., & Kostić, A. (2006). The Balance of Sense Probabilities Affects Processing of Polysemous Words. The 12th Annual Conference on Architectures and Mechanisms for Language Processing. Max Planck Institute for Psycholinguistics, Nijmegen. *Proceedings*, p. 109.
- Filipović Đurđević, D., & Kostić, A. (2008). The effect of polysemy on processing of Serbian nouns. *Psihologija*, 41(1), 69-86.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- 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 Performance*, 22, 1331–1356.
- 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.
- Klepousniotou, E. (2002). The processing of lexical ambiguity: homonymy and polysemy in the mental lexicon. *Brain & Language*, 81(1-3), 205-23.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lund, K., & Burgess, C. (1997). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers*, 28, 203-208.
- McDonald, S. (2000). *Environmental determinants of lexical processing effort*. PhD dissertation, University of Edinburgh, UK.

- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- Medijska dokumentacija Ebart. <http://www.arhiv.co.yu/>.
- Millis, 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 Martín, F. (2006). A Bayesian Information-Theoretical model of lexical processing. The fifth International Conference on the Mental Lexicon, Montreal, Quebec, Canada. *Proceedings*, p. 159.
- Moscoso del Prado Martín, F., Kostić, A., & Filipović Đurđević, D. (submitted). The Bayesian Information-Theoretical Model of Lexical Processing. The Missing Link Between Morphemic Assemblies and Behavioral Responses: a Bayesian Information-Theoretical Model of Language Processing.
- Neal, R. M. (1991). Bayesian mixture modeling by Monte Carlo simulation. *Technical Report CRG-TR-91-2*. Toronto: Department of Computer Science, University of Toronto.
- Neal, R. M. (1998). Markov chain sampling methods for Dirichlet process mixture models. *Technical Report No. 9815*. Toronto: Department. of Statistics, University of Toronto.
- Neal, R. M. (2004). Software for Flexible Bayesian Modeling and Markov Chain Sampling (release of 2004-11-10). Toronto: Department. of Statistics, University of Toronto. <http://www.cs.toronto.edu/~radford/fbm.software.html>.
- Pulvermüller, F. (2003). *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*. Cambridge: Cambridge University Press.
- Pylkkänen, L., Llinas, R., & Murphy, G. (2006). Representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, 18(1), 1–13.
- Rodd, J., Gaskell, M. G., & Marslen-Wilson, W. D. (2001). For better or worse: modelling effects of semantic ambiguity. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 891-896). Mahwah(NJ): Lawrence Erlbaum Associates.
- 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.
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, 28, 89-104.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.
- Schütze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics*, 24(1), 97 – 123.

REZIME

**VEKTORSKI ZASNOVANA SEMANTIČKA ANALIZA
POLISEMIČNIH REČI**

Dušica Filipović Đurđević

Odsek za psihologiju, Univerzitet u Novom Sadu

Đorđe Đurđević

Elektrotehnički fakultet, Univerzitet u Beogradu

Aleksandar Kostić

Laboratorija za eksperimentalnu psihologiju, Univerzitet u Beogradu

Ranija istraživanja pokazala su da je porast u broju povezanih značenja (polisemija) praćen kraćim vremenom obrade reči, dok je porast u broju nepovezanih značenja (homonimija) praćen dužim vremenom obrade reči. U ovom istraživanju evaluirali smo metod za kvantitativno opisivanje višeznačnih reči na osnovu višedimenzionalne distribucije kontekstualnih vektora kojeg su predložili Moscoso del Prado Martín i saradnici. Na osnovu distribucije kontekstualnih vektora pojedinačnih pojavljivanja polisemičnih reči srpskog jezika izračunata je entropija ekvivalentne Gausove distribucije i negentropija funkcije gustine verovatnoće. Entropija ekvivalentne Gausove distribucije, koja je izvedena iz matrice kovarijanse, predstavlja meru varijabilnosti u višedimenzionalnom prostoru i odslikava postojanje povezanih značenja reči (polisemija). Negentropija predstavlja razliku između entropije ekvivalentne Gausove distribucije i diferencijalne entropije funkcije gustine verovatnoće kontekstualnih vektora i odslikava postojanje nepovezanih značenja (homonimija). U skladu sa predviđanjima, na skupu polisemičnih imenica srpskog jezika, zabeležen je samo efekat entropije ekvivalentne Gausove distribucije. Negentropija nije imala uticaj na vreme reakcije. Ovaj nalaz je u skladu sa pretpostavkom da entropija ekvivalentne Gausove distribucije, kao mera širine aktivacije u semantičkom prostoru, odslikava polisemiju, odnosno prisustvo povezanih značenja reči. Stoga se prednost u obradi polisemičnih reči može objasniti širokom aktivacijom u semantičkom prostoru i smanjenom kompeticijom među Gausovim distribucijama koje se preklapaju u velikj meri.

Ključne reči: polisemija, semantički prostor, kontekstualni vektori, diferencijalna entropija, entropija ekvivalentne Gausove distribucije, negentropija

RAD PRIMLJEN: 12.12.2008.