Redesigning the Exploration of Semantic Dynamics – SSD Account in Light of Regression Design

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID	QJE-STD-20-492.R2
Manuscript Type:	Standard Article
Date Submitted by the Author:	25-Aug-2021
Complete List of Authors:	Mišić, Ksenija; University of Belgrade Faculty of Philosophy, Department of Psychology Filipović Đurđević, Dušica; University of Belgrade Faculty of Philosophy, Department of Psychology
Keywords:	semantic ambiguity, Semantic Settling Dynamics model, polysemy, homonymy, entropy, redundancy

DOI: 10.1177/17470218211048386

REDESIGNING THE EXPLORATION OF SEMANTIC DYNAMICS – SSD ACCOUNT IN LIGHT OF REGRESSION DESIGN

Ksenija Mišić[a], Dušica Filipović Đurđević[a, b, c]

[a]*Laboratory for Experimental Psychology, Faculty of Philosophy, University of Belgrade, Serbia*

[b]*Laboratory for Experimental Psychology, Faculty of Philosophy, University of Novi Sad, Serbia*

[c]*Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia*

Correspondence

Ksenija Mišić

ksenija.misic@f.bg.ac.rs

Laboratory for Experimental Psychology, Faculty of Philosophy, University of Belgrade

Čika Ljubina 18-20, 11000 Belgrade, Serbia

tel: +381112630542

DOI: 10.1177/17470218211048386

Redesigning the Exploration of Semantic Dynamics – SSD Account in Light of Regression Design

The Semantic Settling Dynamics model (Armstrong & Plaut, 2016) postulated that the seemingly inconsistent effects of lexical ambiguity are, in fact, a systematic manifestation of the specific dynamics that arise as a consequence of the amount of time spent in processing. The model has thus far been tested by prolonging lexical decision and comparing homonymous, polysemous, and unambiguous words in a factorial design. Here, we kept the strategy of task manipulation but tested the model by using continuous measures as indices of the level of lexical ambiguity and their slopes as indices of the effect size. We expressed the size of the polysemy effect as the slope of the effect of entropy of sense probability distribution and the size of the homonymy effect as the redundancy of sense probability distribution. Comparing lexical decision tasks with the shorter and longer time spent in processing, we observed the predicted decrease in the effect of the polysemy level as well as the predicted increase in the effect of homonymy level.

Keywords: semantic ambiguity, Semantic Settling Dynamics model, polysemy, homonymy, entropy, redundancy

DOI: 10.1177/17470218211048386

Introduction

Lexical ambiguity

A large number of words in various languages can denote multiple referents (Rodd et al., 2004). Intuitively, this seems like a potential problem for understanding language, but native speakers successfully cope with this problem on an everyday basis. Early research typically approached the topic by contrasting ambiguous and unambiguous words. Results dominantly indicated that ambiguous words were processed faster (Rubenstein et al., 1970), although some papers were reporting the absence of ambiguity effect (Borowsky & Masson, 1996; Hino & Lupker, 1996; Rayner & Duffy, 1986). Turning to the linguistic categorization of ambiguity, it was discovered that an additional factor was affecting these results – meaning/sense relatedness. Therefore, lexical ambiguity as the psycholinguistic variable was split into two categories that were further investigated. Words with multiple related senses are called polysemes (e.g., PAPER – a sheet of paper and a scientific paper), whereas words with multiple unrelated meanings are called homonyms (e.g., BANK – a riverbank and a financial institution; Gortan-Premk, 2004; Lyons, 1977; Stanojčić & Popović, 1995). Studies that took this relatedness into account mostly found that the processing advantage in ambiguous words was limited to words with related senses, i.e., polysemous words, whereas unrelated meanings produced a reversed effect and were processed slower compared to unambiguous words (Rodd et al., 2002).

Nonetheless, polysemy and homonymy effects were not consistently observed (for a detailed review, see Eddington & Tokowicz, 2015). The polysemy effect was prominent when studies used tasks with faster response times such as a visual lexical decision task (LDT) or its slower auditory version – auditory LDT (Azuma & Van Orden, 1997; Beretta et al., 2005; Klepousniotou, 2002; Klepousniotou & Baum, 2007; Rodd et al., 2002). When a task with

DOI: 10.1177/17470218211048386

slower response times such as semantic categorization was used, the polysemy effect was absent (Hino et al., 2006). On the other hand, the homonymy effect was present only in auditory LDT (Rodd et al., 2002) and semantic categorization tasks (Hino et al., 2006), but not in visual LDT (Rodd et al., 2002). When embedded in a sentence and preceded by the biasing context, all ambiguous words were mostly slower than unambiguous controls (Brocher et al., 2016; Brocher et al., 2018; Frazier & Rayner, 1990; Rayner & Duffy, 1986).

Semantic Settling Dynamics account

Multiple frameworks were proposed to account for various aspects of lexical ambiguity processing (Hino & Lupker, 1996; Joordens & Besner, 1994; Kawamoto et al., 1994; Rodd et al., 2004, Rodd, 2020). We will focus on one of the recent proposals – the Semantic Settling Dynamics (SSD) account (Armstrong, 2012; Armstrong & Plaut, 2016), which attempted to describe a principle underlying the variety of seemingly opposing findings in lexical ambiguity research. The SSD account assumes that ambiguous words do not differ from unambiguous words, except for the way their meanings/senses are represented. The account postulates that both polysemous and homonymous words are represented via multiple semantic features. Polysemous words share a set of overlapping features (the core), which is a group of features shared by all or most of the senses and a distinctive set of features for every sense. Homonyms, on the other hand, do not share any features and have separate and distant meaning representations.

According to the SSD account, the effects found in the literature arise as a result of the interplay of dominant cooperative processes and weaker competitive processes. Early polysemy processing is dominated by cooperative dynamics – core co-activates all the distinctive features, leading to high activations and large polysemy effects. If the word is

DOI: 10.1177/17470218211048386

processed longer, competitive processes will enter the processing, and although weaker than cooperative processes, they will lead to a reduction in the effect towards the later processing. Conversely, separate features of unrelated homonym meanings lead to a split in the initial activation leading to an effect slowly emerging only towards the end of the isolated processing. The account also describes processing in context. Considering that previous work suggested that context does not have a strong influence during the early processing (Armstrong & Plaut, 2016), we will not dwell into context-sensitive processing.

Essentially, the idea behind the SSD account (Armstrong & Plaut, 2016) is that activation corresponds to the time spent in processing and that the pattern of the dynamics of activation is dependent on the relatedness among the meanings/senses of the ambiguous word. This relatedness affects the cooperative/competitive dynamics and consequently gives rise to different ambiguity effects.

These dynamics correspond to what had previously been seen as the inconsistent empirical effects found throughout ambiguity research. To be exact, strong activation in earlier, cooperation-dominated processing causes a large polysemy effect in the visual and auditory LDT (Klepousniotou, 2002; Klepousniotou & Baum, 2007; Rodd et al., 2002). Later competition among distinctive features of the senses negatively impacts the overall activation, and therefore, polysemy effects are not present in slower tasks, such as semantic categorization (Hino et al., 2002; Hino et al., 2006). Finally, low activation in early homonym processing (due to the division of activation between multiple meanings) is in line with the absence of effect in visual LDT and its late emergence in tasks such as auditory LDT, but also semantic categorization tasks (processing dominated by competition; Hino et al., 2006; Hino et al., 2010; Klepousniotou & Baum, 2007; Rodd et al., 2002).

DOI: 10.1177/17470218211048386

In addition to incorporating familiar findings (mostly related to comparing the processing of polysemy/homonymy to the processing of unambiguous words), the SSD account (Armstrong & Plaut, 2016) also offers clear predictions regarding the relative size of the ambiguity effect. For example, in early processing, a large polysemy effect would be present. It would continue to grow as processing unfolds, only to be attenuated and eventually disappear towards the end of isolated processing. Homonyms, on the other hand, would show no effect in early processing and the effect would slowly emerge after a longer time spent in processing.

The SSD account predictions were previously tested in two ways – by running computational simulations and by collecting empirical data in a series of comparable tasks. The simulations were run in a connectionist model which was rooted in the idea of the PDP model proposed by Rodd et al. (2004) and work by Piercey and Joordens (2000). These results demonstrated that cooperative/competitive dynamics can indeed emerge from such a model and gave strong support to the validity of the account predictions (Armstrong & Plaut, 2016, appendix). Empirical tests were developed with a general strategy of comparing polysemy and homonymy effects to unambiguous words with respect to the different times spent in processing. Polysemy effect was detected early but decreased in late processing. On the contrary, the homonymy effect was absent in early processing but emerged as the processing unfolded.

To minimize qualitative differences in processing, the authors were using different versions of the lexical decision task. Standard visual LDT served as the baseline, whereas multiple modifications were applied in order to prolong the time spent in semantic processing. In three experiments, prolonging was attempted by varying the degree of orthographic similarity between pseudo-words and words (Armstrong, 2012; Armstrong &

DOI: 10.1177/17470218211048386

Plaut, 2016; Medeiros & Armstrong, 2017). Two experiments used stimulus degradation or, more precisely, stimulus-background contrast and visual noise (Armstrong, 2012; Armstrong & Plaut, 2016; Medeiros & Armstrong, 2017). Lastly, three experiments used various auditory manipulations (comparing visual and auditory LDT; auditory noise; and speeding up/slowing down auditory stimuli) in order to achieve prolonged semantic processing (Medeiros & Armstrong, 2017). Although the results of these experiments were mostly in accordance with the model predictions, the evidence was not as strong as expected (Armstrong, 2012; Armstrong & Plaut, 2016, p. 946; Medeiros & Armstrong, 2017). The crucial interactions did not reach significance, making the authors express some concerns for the strength of the evidence. They attributed the observed inconsistencies to methodological issues and some theoretical factors, inviting further tests of the model.

Current goal

Previous testing of the SSD account was based on categorical comparison of polysemy and homonymy effects to the processing of unambiguous words, with manipulating time spent in processing (Armstrong, 2012; Armstrong & Plaut, 2016; Medeiros & Armstrong, 2017). We aimed to test whether time spent in processing would relate to the *size* of the ambiguity effects while applying a *continuous* measure as an index of the ambiguity level (i.e., the predictor of the polysemy and homonymy effects). From a methodological point of view, regression design is more robust to higher standard error values and able to yield higher statistical power for smaller effects (Baayen, 2010). Regression design is shown to be a superior design in psycholinguistic research because it allows using the full variety of words and not artificially creating sets of matched stimuli that might mask the effects and reduce the generalizability of the results (Baayen, 2010; Balling, 2008).

Time-dependent dynamics

To illustrate the strategy, we depicted the predictions of the SSD account (Figure 1) by presenting the predicted size (in absolute values) of the polysemy effect (blue line) and the predicted size (in absolute values) of the homonymy effect (green line). It should be noted that the predicted effects would be in the mutually opposite directions. However, here we want to focus on the dynamics of the effect change, and hence depict their absolute values (thus eliminating the reversed nature of the two effects). We presented the two effects as a function of the time spent in processing (x-axis), during the processing of an isolated word, i.e., before the context is processed. Throughout the literature, ambiguity effects were determined relative to the unambiguous words. Therefore, we also expressed the predicted size of the ambiguity effect as the absolute difference between the predicted activation of the ambiguous word and the predicted activation of the unambiguous word at a given point in time (based on Armstrong 2012, figure 3.1). In other words, each line was obtained by subtracting the predicted activation values of the ambiguous words (polysemous or homonymous) from the predicted activation values of unambiguous words (at a given point in time) and taking the absolute value of the obtained difference in activation. As illustrated in Figure 1, the predicted effect of homonymy increases with time spent in processing, whereas the predicted polysemy effect exhibits very different dynamics. The initial dramatic increase in the size of the predicted polysemy effect is followed by the peak of the effect, which turns into a decline as the processing unfolds. By comparing the two functions (i.e., the sizes of the two effects), we defined two important time windows. First, we identified the area where the polysemy effect is larger than the effect of homonymy (Figure 1, area A). This is expected to be a result of cooperative dynamics between related senses of a polysemous word. Based on previous studies (Armstrong, 2012; Armstrong & Plaut, 2016; Eddington &

DOI: 10.1177/17470218211048386

Tokowicz, 2015), we expected to be able to map this time window onto standard visual LDT, as it had previously been done by the authors of the SSD account. Second, we searched for the area where the processing was prolonged enough for the late emergence of the homonymy effect and the attenuation of the polysemy effect, both being a consequence of competitive dynamics (Figure 1, area B). With this in mind, we set the goal of investigating whether the predicted decrease in the polysemy effect and increase in homonymy effect would be observed when experimentally manipulating the time spent in processing.

--- Insert Figure 1 about here ---

The size of the ambiguity effects

Earlier research expressed the size of the ambiguity effects as the group difference between processing latencies of polysemous or homonymous words relative to unambiguous controls (Armstrong & Plaut, 2016; Hino & Lupker, 1996; Klepousniotou, 2002; Klepousniotou & Baum, 2007; Klepousniotou et al., 2008; Medeiros & Armstrong, 2017; Pexman & Lupker, 1999; Piercey & Joordens, 2000; Rodd et al., 2002). This implied averaging processing times for the groups of words, and thus losing the fine-grained differences within the groups of ambiguous words.

However, not all ambiguous words are equal, as (in addition to the differences in relatedness of meanings/senses) they differ with respect to the number of senses/meanings and the balance of their probabilities, which are separate sources of ambiguity (Filipović Đurđević & Kostić, 2017; Filipović Đurđević, 2019; Gilhooly & Logie, 1980). The more senses/meanings a word has and the more balanced their frequencies of use, the more ambiguous the word is. Put differently, both the number of senses/meanings and the balance of their probabilities are contributing to the uncertainty of the word's true sense/meaning.

DOI: 10.1177/17470218211048386

This uncertainty can be expressed as the information theory measure of entropy ($H = -\Sigma p*\log(p)$, p – sense/meaning probability; Filipović Đurđević & Kostić, 2017; Gilhooly & Logie, 1980). Low values of entropy describe words that are close to unambiguous words and bear little uncertainty of their true sense/meaning (e.g., words with few senses/meanings one of which is highly dominant). High values of entropy are indicative of high sense/meaning uncertainty, i.e., of a very ambiguous word (with many senses/meanings all of which are frequently used). This allows for fine-grained discrimination among the words of different level of ambiguity.

It has been demonstrated that an increase in entropy is followed by a decrease in the processing latencies of polysemous words (Filipović Đurđević, 2007; Filipović Đurđević & Kostić, 2021). Additionally, it has been shown that both the number of senses and the balance of sense probabilities (two contributors to entropy) affect recognition of polysemous words – an increase in the number of senses and an increase in the balance of sense probabilities were followed by a decrease in processing latencies (Filipović Đurđević, 2007; Filipović Đurđević & Kostić, 2021). Balance of probabilities was expressed in terms of information theory measure of redundancy (T), which is calculated based on the ratio of the observed entropy (H) and the maximum entropy (H_{max}) of the probability distribution ($T = 1 - H / H_{max}$; $H_{max} = \log N$; N = number of meanings/senses; Filipović Đurđević & Kostić, 2017; Filipović Đurđević, 2019; Shannon, 1948). Low redundancy is typical for words with balanced sense/meaning probabilities; hence an increase in redundancy was followed by an increase in recognition latencies. In the case of homonyms, the effects of entropy and number of meanings did not reach significance (likely because of the range restriction for the number of meanings for homonyms (Filipović Đurđević, 2019), whereas the balance of meaning probabilities expressed as redundancy did affect recognition latencies – an increase in

DOI: 10.1177/17470218211048386

redundancy (i.e., a decrease in the balance of meaning probabilities) was followed by a decrease in processing latencies. Therefore, within the group of polysemous words, those with many balanced senses were recognized faster, whereas, within the group of homonymous words, those with (many?) balanced meanings were recognized slower. This opposite pattern fits with the SSD account, and this will be discussed further in the general discussion.

In this paper, for each category of ambiguous words, we take the one continuous measure that has been shown to be a good predictor of recognition time, based on previous studies on Serbian polysemous nouns. Our goal was not to demonstrate the superiority of any continuous ambiguity measure over some other continuous measure, but rather to explore whether temporal effect changes are captured by continuous measures, as compared to previously applied categorical comparisons. Therefore, we take the entropy of sense probability distribution as a continuous measure of the level of ambiguity of polysemous words (Filipović Đurđević & Kostić, 2017; Gilhooly & Logie, 1980) and the redundancy of the meaning probability distribution as a continuous measure of the level of ambiguity of homonymous words (Filipović Đurđević, 2019). We will look at the slope of the entropy effect as an indicator of the size of the polysemy effect and the slope of the redundancy effect as an indicator of the size of the homonymy effect. Crucially, we will look at these slopes in two time windows – in the standard lexical decision task (Figure 1, area A) and a version of the same task which is changed to elicit longer processing. By comparing the slopes from the two time windows, we will test the prediction which we derived from the SSD model (Armstrong, 2012; Armstrong & Plaut, 2016). According to the SSD model, the slope of the entropy/redundancy effect should be smaller in the prolonged processing condition for polysemes and larger for homonyms.

DOI: 10.1177/17470218211048386

We present two experiments that tested SSD account predictions in a regression design. The first experiment was aimed at testing the prolonging strategy for polysemous words, and the second one for homonymous words. For the prolonging strategy, we chose the intermodal manipulation (Medeiros & Armstrong, 2017). Visual LDT was aimed to capture early processing (Figure 1, area A) which is also referred to as baseline condition. In order to prolong the processing, an auditory LDT (which is also referred to as prolonged condition) was used to capture the effects in later processing (Figure 1, area B). It was expected that the second task would prolong the processing sufficiently in order to reach the point in time when the effects of polysemy and homonymy were of equal size, i.e., when the two lines in Figure 1 cross. Previous studies indicated that this was observed in auditory LDT (Rodd et al., 2002, Klepousniotou & Baum, 2007). Other research also showed that there are key differences in processing time between these two tasks (Rodd et al., 2002). Namely, the homonymy effect, an effect that requires longer semantic processing in order to emerge, was not present in visual LDT but was significant in auditory LDT. This supports the fact that these two tasks differ in processing time needed for a response and therefore can be used to compare effects between them.

Experiment 1

Materials and methods

Participants

This experiment included a sample of 143 participants, who were first-year psychology students (Department of Psychology, Faculty of Philosophy, University of Belgrade) and some volunteers. All participants signed informed consent forms before taking part in the experiment. The study was approved by the Institutional Review Board of the Faculty of

DOI: 10.1177/17470218211048386

Philosophy of the University of Belgrade. The sample was divided into two groups, one for each experimental condition, so the baseline group consisted of 71 participants and the prolonged of 72 participants. Each participant took part only in one condition, had normal or corrected-to-normal eyesight and their native language was Serbian.

Stimuli and design

Participants were presented with 160 Serbian nouns and 160 pseudo-words. Out of 160 words, 150 were polysemous and were retrieved from the Serbian polysemous words norms (Filipović Đurđević & Kostić, 2017) that contained participant ratings for familiarity and concreteness. The norms also contain the number of senses and sense probabilities estimations from the participant sense production task, from which the Shannon entropy was calculated (Gilhooly & Logie, 1980, Shannon, 1948). Frequency estimations (per 2 million) were taken from the Frequency Dictionary of Serbian language (Kostić, 1999). Control variables were also word length (in letters/phonemes) and orthographic neighbourhood size (Colthart's N; Davelaar et al., 1978). The descriptive statistics of the presented polysemous words are presented in Table 1. The remaining ten words were unambiguous fillers and were not analysed for the purposes of this paper.

Pseudo-words were created in order to be suitable for auditory LDT – only the last syllable was altered (Ernestus & Cutler, 2015), in order for participants to have to listen to the entire duration of the recorded word before making a decision.

The response variable was word processing time (reaction time) which was considered to begin at the onset of the word stimuli and to end with the participant's response.

Stimuli were presented in two conditions of task modality. Each modality was presented to a separate group of participants. The first condition (baseline) was a standard

DOI: 10.1177/17470218211048386

visual LDT, while the second one (prolonged; longer processing in relation to the baseline condition) was an auditory LDT. For the auditory task, stimuli were recorded within a carrier phrase. The carrier phrase was "TA [STIMULUS] PEVA", pronounced as /ta [] peva/ (Eng. *that [STIMULUS] sings*) and each stimulus within this same three-word phrase in order to obtain more natural-sounding stimuli. The target stimulus was then cropped out of the recording and saved as a separate file. Recording and cropping were done in PRAAT 6.0.37 (Boersma & Weenink, 2018). The mean duration of the noun stimuli was 586.21 ms (SD = 61.13 ms).

--- Insert Table 1 about here ---

Procedure

In visual LDT, a trial consisted of a fixation cross (1000 ms), followed by a blank screen (500 ms), and then the stimulus, which remained on the screen until the participant's response or until the 3000 ms time-out. In auditory LDT, a trial consisted of a fixation cross displayed on the screen (1000 ms) to indicate that the auditory stimulus followed. The auditory stimuli were presented over headphones for the duration of each stimulus. After an auditory stimulus was presented, a participant had a maximum of 3000 ms to respond. The procedure was taken from Rodd et al. (2002).

Responses were given by clicking on the left mouse button if the stimulus was a word and the right mouse button if the stimulus was a pseudoword. Participants received feedback if they were too slow or responded incorrectly. The auditory task had the same structure as the visual – two counterbalanced blocks consisting of 80 words (75 polysemous words and five unambiguous fillers) and 80 pseudowords with 20 (10 words, ten pseudowords) practice trials before each block. Practice trial data were not analysed. Tasks were programmed and

DOI: 10.1177/17470218211048386

presented in OpenSesame experimental software (Mathôt et al., 2012).

Results

Data were screened for outliers prior to performing any of the analyses. All participants and words with accuracy lower than 75% were excluded (two words were removed due to low accuracy rates), as were the RTs obtained for incorrect or anticipatory responses (less than 200 ms), amounting to approximately 5% of the original dataset (6% of the visual LDT condition and 5% of the auditory LDT condition). All RTs were transformed using an inverse transformation (-1000/RT; Box & Cox, 1964, as recommended by Baayen & Milin, 2010, Brysbaert, 2017). The order of stimulus presentation for each participant (Trial order) was also controlled for (Baayen & Milin, 2010), and all variables were transformed to z-scores (Gelman & Hill, 2007).

Data were analysed by Generalized Additive Mixed Models using *mgcv* (Wood, 2006, Wood, 2011) and *itsadug* packages (van Rij, Wieling, et al., 2020) in statistical environment *R* (R Core Team, 2018). Model building procedure followed recommendations by Baayen and colleagues (2017)[1].

We employed a mix of confirmatory and exploratory approach. Since the main hypothesis was the ambiguity effect change, the interaction between modality and index of ambiguity size index (in this particular experiment – entropy) was included in every model. We departed from a fully confirmatory approach for a variety of reasons, the main reason being across ambiguity literature, exploring interactions with other lexical and semantic variables was found to be an important issue (Gernsbacher, 1984; Jager et al., 2016; Jager &

[1] Step by step process of variable and interaction selection is available in supplementary data.

DOI: 10.1177/17470218211048386

Cleland, 2016; Medeiros & Armstrong, 2017; Rice et al., 2019). Furthermore, generalized additive mixed models allow for testing the complex non-linear interactions that allow us a more nuanced insight into our data. It allows non-linearities, but no prior assumptions on the shape of a curve or a surface need to be made (van Rij, Vaci, et al., 2020). The final important reason for using the exploratory approach was to be able to create a parsimonious model (Bates et al., 2015) that fits the data the best and does not overfit. Decisions on whether to include a parameter were made by comparing nested models with only the parameter of interest being excluded. To summarize – starting from the theory we predicted entropy by task modality interaction which was the main focus of our analysis. In addition to this interaction, we also tested for the effects of multiple psycholinguistic variables along with their interactions with our variables of interest. We did so by following two goals – to introduce additional control and to add to the body of knowledge motivated by the already existing interest in the various interactions (Rice et al., 2019).

The final model included (1) trial order, word length (in letters; 2), log-transformed word frequency (3) and orthographic neighbourhood size (4). The model also contained the task modality (5) as a fixed factor (visual/auditory LDT, reference level being the visual LDT condition) in order to explicitly test the prolonging of the processing. For the manipulation factor, the reference level was the visual LDT condition. The interaction between entropy and familiarity in two modality conditions – visual (6a) /auditory (6b) manipulation was modelled by tensor product smooths. Random effects included: by-participant factor smooths for trial order (7), random slopes for word length (8), and by-item random slopes for task modality manipulation (9). The model that is presented here was the refitted model, after the extreme residual removal (Table 2). Model refitting led to removing an additional 3% of the dataset.

--- Insert Table 2 about here ---

DOI: 10.1177/17470218211048386

The control predictors' effects were in accordance with the majority of previously reported findings. As presented in Table 2 (parametric coefficient), the word length effect was inhibitory, the word frequency effect was facilitatory, and surprisingly, orthographic neighbourhood size was inhibitory. The trial order showed fatigue effects throughout the experiment. The effect of the task modality manipulation was significant, and processing was considerably longer in the prolonged condition (auditory LDT).

Our approach to testing interactions started from the interaction of the main theoretical interest, the two-way interaction between task modality and entropy. In order to explore the possible interactions of ambiguity and other lexical and semantic variables we tested several three-way interactions by adding other lexical/semantic predictors to the aforementioned two-way interaction. Each of the three-way interactions was compared against a model with two two-way interactions (entropy by task modality and one of the lexical/semantic predictors by entropy[2]). These interactions were tested because of the previous findings that many of the lexical/semantic variables used as control variables in this design, modulate the ambiguity effects (Rice et al., 2019)

Out of all possible three-way interactions the one which was best supported by the data was the three-way interaction between entropy and familiarity by task modality. This interaction had the lowest AIC value (see supplementary materials for more details) compared to the model with two two-way interactions and compared to other three-way interaction models

[2] The third possible two-way interaction, between one of the lexical/semantic predictors and task modality, was not included as theoretical implications of such interactions are unknown.

DOI: 10.1177/17470218211048386

Furthermore, we tested adding additional interactions to the model including the entropy and familiarity by task modality interaction, however model comparisons revealed that such addition was not supported by the data. Therefore, to avoid overfitting, we did not include any other three- or two-way interactions in that model.

Finally, the model containing only the three-way interaction between task modality, entropy and familiarity was revealed to be the most suitable (tensor product smooths; Table 2, Lines 6a and 6b). This three-way interaction indicated different patterns of entropy by familiarity interaction in the two tasks. In the baseline condition, familiarity and entropy cooperated and with the increase in both measures, the processing latency decreased (visual LDT; Figure 2). However, only the effect of familiarity was present in ALD – no entropy effect at any value of familiarity was found in prolonged condition (auditory LDT; Figure 2). Importantly, none of the investigated control variables and their interactions affected the interaction of main interest (entropy by task modality interaction).

--- Insert Figure 2 about here ---

Discussion

Our first experiment managed to capture the two points on the account's temporal dimension, by comparing auditory and visual LDT. The RT difference between the visual and auditory LDT was significant and enabled us to draw conclusions about SSD account predictions from these results. The interaction confirms our hypothesis by demonstrating that if we prolong the semantic processing, the regression slope of entropy will reduce. In other words, polysemy effect size would change as is presented in differences between the blue line in area A and area B in Figure 1. However, in our experiment, the effect completely

DOI: 10.1177/17470218211048386

disappeared. Although our prediction was that in the prolonged condition, we would detect a reduction, and not the absence of the polysemy effect (blue line in Figure 1, area B), this finding still falls in line with the account predictions, albeit later on the temporal dimension of the model. The lack of the polysemy effect in the auditory LDT was also detected when the same manipulation was used to test the account (Medeiros & Armstrong, 2017). Such a finding leads to the conclusion that our task modality manipulation was more difficult and prolonged the processing further than we anticipated. The interaction between entropy and familiarity amounts to the existing evidence that highly familiar words are processed equally fast regardless of being more or less ambiguous.

Additionally, we detected an inhibitory effect of orthographic neighbourhood size, which is contrary to the dominant findings. Considering the theoretical focus of the paper, and the status of the orthographic neighbourhood size as a control variable in our design, we will not discuss this in great detail. However, we believe that this finding is related to the highly transparent orthography of Serbian language and the fact that the orthographic neighbourhood is simultaneously a phonological neighbourhood. Previous research has typically documented inhibitory effects of phonological neighbourhood during auditory word presentation (Goh et al., 2009; Luce & Pisoni, 1998; Ziegler et al., 2003), and sometimes even in visual lexical decision task (Yap & Balota, 2009; Yates, 2005; Yates et al., 2004). This is further reinforced by half of our data being presented in auditory modality. The possibility of the confounding with phonological neighbourhood effect may be explicitly tested by introducing interaction between the ONS and modality. This interaction was added to the fixed effect part of the model. Comparisons to the reported model revealed that this addition does not improve the fit, and therefore is not suitable to be left in the model. Although it is advised against checking the effects (and their probabilities) in any model except the final one

DOI: 10.1177/17470218211048386

(Baayen et al., 2017), we checked the model summary. This gave us further evidence that we may have captured phonological neighbourhood size effects in this case, since the ONS effect was present only in the auditory LDT, whereas it was absent in visual task.

Experiment 2

Materials and methods

Participants

The sample consisted of 141 first-year students from the Department of Psychology, Faculty of Philosophy, University of Belgrade. All participants were native speakers of Serbian and had a normal or corrected-to-normal vision. They were randomly assigned into one of two experimental conditions; the first one being the baseline group (visual LDT condition) consisting of 72 participants; and the second one being the experimental group (auditory LDT-condition) consisting of 69 participants. Before taking part in the experiment, participants signed an informed consent form. The study was approved by the Institutional Review Board of the Faculty of Philosophy of the University of Belgrade. No participant from experiment 1 entered the sample for this experiment.

Stimuli and design

The number of senses, the entropy of meaning probability distribution, and redundancy effects (ambiguity measures) were tested in two between-participant conditions – visual LDT and auditory LDT (task modality condition). Control variables were familiarity, concreteness, word frequency (per 2 million; Kostić, 1999), word length (in letters) and orthographic neighbourhood size (Colthart's N; Davelaar et al., 1978).

DOI: 10.1177/17470218211048386

The number of meanings and their probabilities, from which entropy and redundancy of the meaning probability distribution are calculated, were obtained in a norming study, as well as familiarity and concreteness ratings. The procedure for the norming study was taken from Filipović Đurđević (2019). A separate group of 30 participants listed meanings of 54 homonymous nouns in a meaning production task, following the total meaning metric approach (Azuma, 1996). The initial list of homonyms was selected according to the dictionary criterion (Rodd et al., 2002), according to which the same letter string listed in two or more separate entries is considered a homonym. Both entries had to have the same orthographic and phonological form (e.g., stress). Stimuli were taken from the Matica Srpska Dictionary of Serbo-Croatian literary language unabridged (1967 – 1974) and abridged (2007) dictionaries.

The number of meanings was estimated by counting every meaning of a word that was listed by at least two participants. Estimation of probabilities was done by calculating the frequency of occurrence for each of the meanings within our sample. Entropy and redundancy were calculated relying on those estimated probability values following Filipović Đurđević (2019). The same group of participants also rated the familiarity and concreteness of the words on a 7-point Likert scale (for results of the norming study, see Table 3).

After the norming study, two words from the initial homonym list were excluded from the analyses. For one, only one meaning was listed by participants and the other one for stress differences between meanings.

In addition to 52 normed homonyms, participants were presented with 20 unambiguous nouns as fillers and 73 pseudowords in the main experiment. Pseudowords were created by changing one letter in the last syllable so the participants would not be able to recognize pseudowords at the onset of the auditory stimulus. With the exception of 52 homonyms, no other stimuli were analysed for the purposes of this paper. The auditory

DOI: 10.1177/17470218211048386

stimuli recording procedure was the same as for experiment 1. The mean duration of the homonymous noun stimuli was 523.83 ms (*SD* = 94.79 ms).

--- Insert Table 3 about here ---

Procedure

The procedure for this experiment for both tasks was the same as in experiment 1. The experiment had the main experimental block consisting of 72 words and 73 pseudowords and a practice block consisting of six words and six pseudowords. Practice trials were not analysed for the purposes of this paper.

Results

Screening for outliers and variable transformations were conducted in the same way as in Experiment 1. Participants and stimuli with accuracy lower than 75% were removed from the dataset. This led to the removal of data from two participants, one from each experimental condition and seven homonymous nouns. Reaction times on incorrect responses, and the ones shorter than 200 ms (anticipatory responses) were also removed from the dataset. Approximately 21% of the original dataset, 21% from the visual, and 20% from the auditory LDT condition was removed.

Data were analysed using Generalized Additive Mixed Models (Wood, 2006; 2011). The model building followed recommendations by Baayen, and colleagues (2017). Decision rules for including parameters in the model are laid out in detail in the results section of Experiment 1. Model comparison procedure was used for selecting fixed effect parameters and in the final model, trial order (1), word length (2), and task modality (with visual LDT set as the reference level; 3) were included. The interaction among familiarity, redundancy, and

DOI: 10.1177/17470218211048386

the task modality was modelled by tensor product smooths (visual LDT: 4a, auditory LDT: 4b). The model also included the autocorrelation parameter. Random effect structure consisted of by-participant factor smooths for trial order (5) and by-item task modality random slopes (6). After we removed extreme residuals, we refitted the model and results from that model are presented in Table 4. This removed an additional 3% of data from the initial dataset.

--- Insert Table 4 about here ---

As expected, the word length effect was inhibitory, and the trial order revealed fatigue effects as the experiment progressed. The manipulation was successful in the substantial prolonging of the processing time in the prolonged condition.

The approach to modelling interactions was the same as in experiment 1. We selected an interaction that was best supported by the data. Again, the initial model contained the entropy by modality interaction and then we proceeded to add other lexical/semantic variables to test whether they modulate the effect. The three-way interaction between modality, redundancy, and familiarity was significant (tensor product smooths; Table 4, 4a and 4b). and best supported by the data. It revealed a different pattern of results for the two tasks. In the visual LDT (Figure 3) there was no redundancy effect, only the familiarity effect. On the contrary, in the auditory task (Figure 3), redundancy cooperated with familiarity and with the increase in redundancy (i.e., decrease in the balance of meaning probabilities), the processing became faster.

Unlike the first experiment, the main interaction of interest was not present unless familiarity was included in the three-way interaction. This is congruent with previous findings where redundancy effect for homonyms was revealed only when familiarity was present in the interaction (Filipović Đurđević, 2019). This will be discussed further in the next section.

DOI: 10.1177/17470218211048386

--- Insert Figure 3 about here ---

Discussion

The second experiment also captured two points on the SSD account's temporal dimension. The RT difference between visual and auditory tasks confirms this and allows us to make conclusions about the possible effect changes. Significant diagnostic interaction between redundancy, familiarity, and manipulation condition shows that the homonymy effect emerges only in prolonged processing, although in cooperation with familiarity. In the baseline condition, no homonymy effect was found, only familiarity effect. This pattern of results is congruent to our predictions regarding the homonymy effect size presented as a difference between the green line in area A and area B in Figure 1. The same results were obtained in the original use of auditory/visual manipulation (Medeiros & Armstrong, 2017).

As in Experiment 1, familiarity was included in a three-way interaction with entropy and modality. Similar to the findings of Filipović Đurđević (2019) continuous measures of homonymy do not seem to emerge unless familiarity is introduced in the interaction. There may be multiple reasons for this. Generally, homonymy effects are weaker and harder to detect (Eddington & Tokowicz, 2015), even in the prolonged conditions such as ours (Armstrong & Plaut, 2016; Medeiros & Armstrong, 2017).

In addition to theoretical reasons, comparing all other three-way interactions revealed a lower AIC value for the model that included familiarity. Unlike in Experiment 1, this difference was not as persuasive, but other indices of fit also supported the inclusion of familiarity. Findings from this experiment are perfectly aligned with the SSD model predictions, with the addition of the familiarity interaction. The by-familiarity interaction revealed that highly familiar homonyms are processed equally fast regardless of their

DOI: 10.1177/17470218211048386

redundancy. This finding mimics the by-familiarity interaction observed in Experiment 1 showing that highly familiar polysemous words are not sensitive to the level of ambiguity as expressed by entropy of sense probabilities.

Once again, we detected an inhibitory effect of orthographic neighbourhood size, which furthers our interpretation of this effect given in the discussion of Experiment 1.

General discussion

The aim of this paper was to test the SSD account (Armstrong & Plaut, 2016) predictions regarding the changes in the ambiguity effects in relation to the time spent in processing. Following the previous practices in the testing of the SSD account, we took both polysemy and homonymy as our test cases. However, unlike the previous testing, we described lexical ambiguity by using the *continuous measure* and tested the effect change in a *regression design*. Our goal was not to compare the explanatory power of measures but to demonstrate the predicted temporal effect differences. Therefore, for the continuous measure, for each of the categories of ambiguous words, we chose one quantification that was previously found to be a good predictor of processing latencies of Serbian polysemous nouns. For polysemy, we chose entropy, a measure of sense uncertainty (Filipović Đurđević, 2007; Gilhooly & Logie, 1980) and for homonymy, we chose redundancy, a measure of the balance of meaning probabilities (Filipović Đurđević, 2019). Based on the SSD account (Armstrong & Plaut, 2016), we derived the prediction that the slope of the entropy effect would weaken with the prolonged processing of polysemous words. For homonyms, we derived the opposite prediction – that the slope of the redundancy effect would become stronger in the prolonged processing. In experiment 1, we observed a significant effect of

DOI: 10.1177/17470218211048386

sense uncertainty (the entropy of sense probability distribution) in visual LDT and the absence of this effect in auditory LDT. In experiment 2, we observed the effect of the balance of meaning probabilities (the redundancy of meaning probability distribution, experiment 2) in auditory LDT and its absence in visual LDT. Overall, our results seem to be in accordance with the predictions of the SSD account we derived.

Experiment 1 delayed the processing enough to observe the attenuation of the entropy effect. Our goal was to administer auditory LDT in order to prolong the processing as compared to visual LDT (following Medeiros & Armstrong, 2017). The manipulation was successful, as overall processing time was significantly longer in the auditory LDT condition. Consequently, the observed reduction of entropy effect, compared to the baseline (as observed in visual LDT), was in accordance with the predictions derived from the SSD model.

However, there are some discrepancies with the existing literature regarding polysemy processing in auditory LDT. As stated in Medeiros and Armstrong (2017), the choice of this manipulation was motivated by the different patterns of results between visual and auditory tasks in Rodd et al. (2002). In work by Rodd and colleagues, a polysemy advantage was present in both tasks, same as in experiments by Klepousniotou and Baum (2007). Medeiros and Armstrong did not find a polysemy advantage in an auditory LDT when it was used as manipulation in order to increase the amount of semantic processing. We seem to replicate this lack of polysemy effect, even when polysemous words are described by entropy. According to the SSD account, this suggests that auditory manipulations were sufficiently difficult, and were able to push the processing in the zone where the polysemy effect is diminished (Figure 1, blue line in area B). Reaching such a late moment in the processing, where competitive dynamics between distinctive features halt the increase in the activation of the polysemous words, was previously thought to be difficult to reach with a lexical

DOI: 10.1177/17470218211048386

decision task. Our finding from the first experiment offers additional evidence that LDT is able to tap into later processing involving competitive dynamics.

As part of the design change, we have introduced continuous measures as an index of the ambiguity level. For polysemy, we used the entropy of the probability distribution (Filipović Đurđević & Kostić, 2017; Gilhooly & Logie, 1980) which is a measure combining two important properties of polysemy – number of senses and balance of their probabilities. Both measures come from information theory, and such a choice was made for comparability between our experiments. Previously, most research on polysemous and homonymous words was not using comparable measures. Polysemous words were mainly described in the sense relatedness terms (metaphor/metonymy) whereas homonyms were mostly described by the balance of their sense probabilities (e.g., Klepousniotou et al., 2008). Entropy and redundancy allow us to compare between experiments, as well as to compare with previous research, also employing regression design (Filipović Đurđević & Kostić, 2017; 2021; Filipović Đurđević, 2019).

According to the SSD account, early processing is dominated by cooperation among different senses, so the processing in that time window is affected by words having balanced probabilities and many senses. Our findings seem to confirm this, along with previous research (Filipović Đurđević, 2007; Filipović Đurđević & Kostić, 2021). This suggests that more uncertainty (i.e., more senses and more balance) leads to a larger overall activation of the polysemous word representations. Less uncertain senses, i.e., less ambiguous polysemous words, have a lower initial activation, and therefore exhibit a smaller effect. In other words, low entropy polysemes are processed more like unambiguous words (Armstrong et al., 2012). This further suggests that there is a whole continuum of different probability distributions between the trajectories of activation for polysemes and unambiguous words. Work on

DOI: 10.1177/17470218211048386

meaning/sense dominance supports this by finding differing patterns of results for priming dominant and subordinate meanings/senses (Klepousniotou et al., 2008; Klepousniotou et al., 2012; Maciejewski & Klepousniotou, 2020).

To further test predictions of the SSD account, we conducted our second experiment where we applied the same manipulation as in our first experiment to the processing of the homonymous words. The results give further evidence to theoretical claims about ambiguity effects changing as a function of time spent in processing. In our baseline condition, we did not detect any ambiguity effect. Only when processing was prolonged, did the homonymy effect emerge. These results, along with the results from experiment 1, go in line with those of Medeiros and Armstrong (2017). Although we did not replicate the findings from the original use of this manipulation (Medeiros & Armstrong, 2017), we did observe a decrease in polysemy effect and an increase in homonymy effect with an increase in time spent in processing.

Our second experiment adds to the body of work that successfully detected the homonymy effect (Armstrong & Plaut, 2016; Armstrong et al., 2012; Beretta et al., 2005; Filipović Đurđević, 2019; Hino et al., 2002; 2006; Medeiros & Armstrong, 2017; Rodd et al., 2002). The lack of homonymy effect in some research (Klepousniotou, 2002; Klepousniotou & Baum, 2007) could be attributed to the processing not being long enough for effect to emerge. Following Filipović Đurđević (2019) we opted to use redundancy of the meaning probability distribution as it was previously successful in predicting latencies, unlike entropy and number of meanings. Redundancy represents a deviation from maximal entropy (where all possible outcomes are equally probable). So, low redundancy words are words that have equal probabilities for each of the meanings, and as redundancy increases, so does reaction time decrease.

DOI: 10.1177/17470218211048386

The SSD account claims that homonym processing is dominated by early competitive processing. Redundancy (balanced/unbalanced meaning frequencies) affects competitive processing (Armstrong & Plaut, 2016, p. 245; Yurchenko et al., 2020; Maciejewski & Klepousniotou, 2020). Initially in processing, activation is split between meanings and cannot reach sufficient activation for effect to be detected. Only in later processing, as competitive processes have influenced the redistribution of activation, the homonymy effect is found. Similarly to how entropy refines the space between "ideal" polysemes and unambiguous words, redundancy can do that partially for homonyms. Two equally probable meanings will have to rely more on competitive processing and delay the disambiguation further. For words where there is a dominant meaning, activation will initially be more biased towards the more probable meaning. This leads again to high redundancy words (with large probability imbalance) being processed more like unambiguous words compared to homonyms (Armstrong et al., 2012, Maciejewski & Klepousniotou, 2020).

We should consider our findings along with the ones observed by Armstrong and colleagues (2012), where they described homonyms using a similar and highly correlated measure. They relied on the measure that represents the frequency of the most probable meaning (β, "biggest"; also similar to the dominance score introduced by Rodd et al., 2002). Although highly similar to redundancy, this measure does not describe the entire distribution of probabilities. Nevertheless, both measures demonstrate preference in processing (smaller recognition latencies) for unbalanced homonyms, which then rely less on the competitive processing, and more on the initial state of the activation. Additionally, both redundancy and β outperform entropy in predicting lexical decision for homonyms (Armstrong et al., 2012; Filipović Đurđević, 2019). These findings highlight that meaning probabilities are of high importance to homonym processing because, with the addition of our results, some

DOI: 10.1177/17470218211048386

description of probability distributions leads to a more reliable path to detecting the homonymy effect.

Another issue that we may consider is the nature of the manipulation used to prolong processing time. According to previous research, including papers pertaining to SSD account, differences in processing between visual and auditory LDT were found (Armstrong & Plaut, 2016; Eddington & Tokowicz, 2015; Medeiros & Armstrong, 2017; Rodd et al., 2004). Throughout the SSD account literature (Armstrong, 2012; Armstrong & Plaut, 2016; Medeiros & Armstrong, 2017), it is stated that the effects observed with homonyms were a crucial test of predictions regarding isolated word processing. In a fast task, such as lexical decision, polysemy effect was rarely an issue – it seemed to emerge in many variations of the task. On the other hand, the homonymy effect was slightly more elusive and task dependent. In the extensive literature review, one of the points made was exactly this seeming irregularity of the homonymy effect (Eddington & Tokowicz, 2015). Rodd and colleagues (2002) found a homonymy effect in the auditory LDT, which was the motivation for using auditory/visual manipulation in the model testing – two tasks seemed to encompass different effect states in the SSD account (figure 1; areas A and B). Another motive was the previous use of such manipulation in testing the SSD account (Medeiros & Armstrong, 2017).

However, relying on differences in processing between the two modalities may raise some concerns. The different nature of the processing in two modalities may lead to different influences of the same effects in two tasks. For example, one possible issue might be a cohort effect (Wurm et al., 2006) that is present in the auditory processing of words and not in visual processing. However, the SSD account does not exclude other effects (such as interactions with other linguistic variables and modality differences); it simply does not describe these effects (yet). What it does describe are purely semantic effects in relation to the activations

DOI: 10.1177/17470218211048386

of the word sense/meaning representation. An alternative explanation of our results would need to somehow account for the inverse effect of some modality-specific variable on polysemy and homonymy. For example, in our case, the modality did not affect the modulatory effect of familiarity.

In addition to changing ambiguity effects, another prominent finding was the observation of interactions between ambiguity measures and familiarity. Although our study was not designed to explicitly test for this interaction, the advanced statistical technique which we applied enabled us to test for the many complex interactions important for lexical ambiguity processing (Rice et al., 2019). The one interaction which was justified by the data[3] was the familiarity (subjective frequency) by ambiguity interaction. Familiarity affected the processing of the ambiguous words in such a way that the effect of ambiguity was less prominent for highly familiar words. This was true in all cases where we found any ambiguity effect (Baseline condition of Experiment 1, and prolonged processing condition of Experiment 2), as well as in previous research (Filipović Đurđević, 2007; Filipović Đurđević, 2019, Filipović Đurđević & Kostić, 2021). Our findings are in accordance with the similar previously reported interactions in which the increase in the frequency of polysemous words reduced the polysemy effect (Jager et al., 2016; Rice et al., 2019). However, the frequency by ambiguity interaction was not found for homonyms in other research (Medeiros & Armstrong, 2017).

The exploration of interactions of ambiguity measures with different lexical and semantic variables is as old as the exploration of ambiguity effects itself. One of the earliest and most prominent findings was the finding that experiential familiarity (Gernsbacher, 1984) is the only reliable predictor, and that ambiguity effects from previous literature stem from

[3] See supplementary materials for model comparisons

DOI: 10.1177/17470218211048386

not controlling for familiarity. Later research discovered that polysemy advantage was present only in low (printed) frequency words, whereas in high-frequency words the effect was reduced or even reversed (Jager et al. 2016; Rice et al., 2019). However, this is not a completely consistent finding considering some studies found no frequency interaction (Medeiros & Armstrong, 2017)

The frequency by ambiguity interaction is typically discussed in terms of feedback activation from the semantic level to the level of orthography (Hino & Lupker, 1996). This is further used to account for ambiguity effects by using the dynamics related to the decision process and dismiss the dynamics within the semantic level as the driving force of the lexical ambiguity effects (Hino et al., 2006). However, although the SSD account (Armstrong & Plaut, 2016) does not explicitly consider such interactions, they are not incompatible with it. Importantly, the SSD account and the decision system account are not considered mutually exclusive. Therefore, our finding of familiarity-by-ambiguity interaction would not dismiss the SSD account but would point to the need for further elaboration. At the same time, our main finding of the lexical-ambiguity-type-specific changes in ambiguity effects as imposed by the prolonged processing seem to fit better with the dynamics at the semantic level, i.e., with the predictions of the SSD account.

Supplementary material

Data and code associated with this article can be found at https://osf.io/wntg9/ .

DOI: 10.1177/17470218211048386

References

Armstrong, B. C. (2012). *The Temporal Dynamics of Word Comprehension and Response Selection: Computational and Behavioral Studies*. [Doctoral dissertation, Carnegie Mellon University]. http://blairarmstrong.net/papers/Armstrong.12.DISSERTATION.html

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966. https://doi.org/10.1080/23273798.2016.1171366

Azuma, T. (1996). Familiarity and relatedness of word meanings: Ratings for 110 homographs. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 109-124. https://doi.org/10.3758/BF03203645

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 Language*, *36*(4), 484–504. https://doi.org/10.1006/jmla.1997.2502

Baayen, R. H. (2010). A real experiment is a factorial experiment? *Mental Lexicon*, *5*(1), 149–157. https://doi.org/10.1075/ml.5.1.06baa

Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.1287/mksc.12.4.395

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. Journal of Memory and Language, 94, 206-234. https://doi.org/10.1016/j.jml.2016.11.006

DOI: 10.1177/17470218211048386

Balling, L. (2008). A brief introduction to regression designs and mixed-effects. *Copenhagen Studies in Language*, *36*(2008), 175–192. http://forskningsbasen.deff.dk/View.external?recordId=auau:12708886

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. *arXiv:1506.04967* [stat.ME].

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Brain Research. Cognitive Brain Research*, *24*(1), 57–65. https://doi.org/10.1016/j.cogbrainres.2004.12.006

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 63–85. https://doi.org/10.1037/0278-7393.22.1.63

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252. Retrieved from http://www.jstor.org/stable/2984418

Brocher, A., Foraker, S., & Koenig, J.-P. (2016). Processing of irregular polyemes in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1798-1813. https://doi.org/10.1037/xlm0000271

DOI: 10.1177/17470218211048386

Brocher, A., Koenig, J.P., Mauner, G., & Foraker, S. (2018). About sharing and commitment: the retrieval of biased and balanced irregular polysemes, *Language, Cognition and Neuroscience*, *33*(4), 443-466. https://doi.org/10.1080/23273798.2017.1381748

Brysbaert, M. (2017). Age of acquisition ratings score better on criterion validity than frequency trajectory or ratings "corrected" for frequency. *Quarterly Journal of Experimental Psychology*, *70*(7), 1129–1139. https://doi.org/10.1080/17470218.2016.1172097

Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: the current state of the literature. *Psychonomic Bulletin & Review*, *22*, 13–37. https://doi.org/10.3758/s13423-014-0665-7

Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decisions. *Quarterly Journal of Experimental Psychology*, 68(8), 1469-1488. https://doi.org/10.1080/17470218.2014.984730

Filipović Đurđević, D. (2007). *Efekat višeznačnosti na obradu izolovanih reči u srpskom jeziku*. [The ambiguity effect in processing of isolated words in Serbian] [Doctoral dissertation, University of Belgrade].

Filipović Đurđević, D., & Kostić, A. (2021). We probably sense sense probabilities. *Language, Cognition and Neuroscience*, 1-28. https://doi.org/10.1080/23273798.2021.1909083

Filipović Đurđević, D., & Kostić, A. (2017). Number, Relative Frequency, Entropy, Redundancy, Familiarity, and Concreteness of Word Senses: Ratings for 150 Serbian

DOI: 10.1177/17470218211048386

Polysemous Nouns. In S. Halupka-Rešetar & S. Martinez-Ferreiro (Eds.), *Studies in Language and Mind* (pp. 13–77). Novi Sad: Filozofski fakultet u Novom Sadu. http://digitalna.ff.uns.ac.rs/sadrzaj/2017/978-86-6065-446-7

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*(2), 181–200. https://doi.org/10.1016/0749-596X(90)90071-7

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Gernsbacher, M. A. (1984). Resolving 20 Years of Inconsistent Interactions Between Lexical Familiarity and Orthography, Concreteness, and Polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*(4), 395–427. https://doi.org/10.3758/BF03201693

Goh, W. D., Suárez, L., Yap, M. J., & Tan, S. H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. *Psychonomic Bulletin & Review*, *16*(5), 882-887. https://doi.org/10.3758/PBR.16.5.882

DOI: 10.1177/17470218211048386

Gortan-Premk, D. (2004). *Polisemija i organizacija leksičkog sistema u srpskome jeziku*. [*Polysemy and organization of the lexical system*]. Zavod za udžbenike i nastavna sredstva.

Hino, Y., Kusunose, Y., & Lupker, S. J. (2010). The relatedness-of-meaning effect for ambiguous words in lexical-decision tasks: when does relatedness matter? *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, *64*(3), 180–196. https://doi.org/10.1037/a0020475

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1331–1356. https://doi.org/10.1037/0096-1523.22.6.1331

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language*, *55*(2), 247–273. https://doi.org/10.1016/j.jml.2006.04.001

Joordens, S., & Besner, D. (1994). When Banking on Meaning Is Not (Yet) Money in the Bank: Explorations in Connectionist Modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1051–1062. https://doi.org/10.1037/0278-7393.20.5.1051

Kawamoto, A. H., Farrar, W. T., & Kello, C. T. (1994). When Two Meanings Are Better Than One: Modeling the Ambiguity Advantage Using a Recurrent Distributed Network. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1233–1247. https://doi.org/10.1037/0096-1523.20.6.1233

DOI: 10.1177/17470218211048386

Klepousniotou, E. (2002). The Processing of Lexical Ambiguity: Homonymy and Polysemy in the Mental Lexicon. *Brain and Language*, *81*(1–3), 205–223. https://doi.org/10.1006/brln.2001.2518

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*(1), 1–24. https://doi.org/10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making Sense of Word Senses: The Comprehension of Polysemy Depends on Sense Overlap. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1534–1543. https://doi.org/10.1037/a0013012

Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language*, *123*(1), 11–21. https://doi.org/10.1016/j.bandl.2012.06.007

Maciejewski, G., & Klepousniotou, E. (2020). Disambiguating the ambiguity disadvantage effect: Behavioral and electrophysiological evidence for semantic competition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000842.supp

Kostić, Đ. (1999). *Frekvencijski rečnik savremenog srpskog jezika* [*Frequency Dictionary of Contemporary Serbian Language*]. Institut za eksperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju

DOI: 10.1177/17470218211048386

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Lyons, J. (1977). *Semantics. Volume I.* Cambridge. University Press. https://doi.org/10.1017/CBO9781139165693

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

Medeiros, J., & Armstrong, B. C. (2017). Semantic Ambiguity Effects: A Matter of Time? In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates.

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology*, *53*(4), 323–334. https://doi.org/10.1037/h0087320

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning Memory and Cognition, 30*(6), 1252–1270. https://doi.org/10.1037/0278-7393.30.6.1252

Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory and Cognition*, *28*(4), 657–666. https://doi.org/10.3758/BF03201255

DOI: 10.1177/17470218211048386

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rice, C. A., Tokowicz, N., Fraundorf, S. H., & Liburd, T. L. (2019). The complex interactions of context availability, polysemy, word frequency, and orthographic variables during lexical processing. *Memory and Cognition*, *47*(7), 1297–1313. https://doi.org/10.3758/s13421-019-00934-4

RMS – Rečnik Matice Srpske. (1967 – 1976). *Rečnik srpskohrvatskoga književnog jezika, t. I-VI* [*Dictionary of Serbo-Croatian literary language*]. Matica Srpska.

RSJ – Rečnik Srpskoga Jezika. (2007). *Rečnik srpskohrvatskoga književnog jezika,* [*Dictionary of Serbo-Croatian literary language*]. Matica Srpska.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language*, *46*(2), 245–266. https://doi.org/10.1006/jmla.2001.2810

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, *28*(1), 89–104. https://doi.org/10.1016/j.cogsci.2003.08.002

DOI: 10.1177/17470218211048386

Rodd, J. M. (2020). Settling Into Semantic Space: An Ambiguity-Focused Account of Word-Meaning Access. *Perspectives on Psychological Science*, *15*(2), 411-427. https://doi.org/10.1177/1745691619885860

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 487–494. https://doi.org/10.1016/S0022-5371(70)80091-3

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Stanojčić, Ž., & Popović, L. (1995). *Gramatika srpskoga jezika* [*Grammar of the Serbian language*](Fourth ed.). Zavod za udžbenike i nastavna sredstva.

van Rij, J., Vaci, N., Wurm. L., & Feldman L. B. (2020). Alternative quantitative methods in psycholinguistics: Implications for theory and design. In V. Pirrelli, I. Plag, & W. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon* (pp. 83-126). De Gruyter. https://doi.org/10.1515/9783110440577-003

van Rij, J., Wieling M., Baayen R. H., & van Rijn, H. (2020). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. R package version 2.4. https://cran.r-project.org/web/packages/itsadug.

Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, *48*(4), 445-464. https://doi.org/10.1111/j.1467-842X.2006.00450.x

DOI: 10.1177/17470218211048386

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

Wurm, L., Ernestus, M., Schreuder, R., & Baayen, R. H. (2006). Dynamics of the auditory comprehension of prefixed words: Cohort entropies and Conditional Root Uniqueness Points. *The Mental Lexicon, 1*(1), 125–146. https://doi.org/10.1075/gest.8.3.02str

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502-529. https://doi.org/10.1016/j.jml.2009.02.001

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1385–1397. https://doi.org/10.1037/0278-7393.31.6.1385

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11*(3), 452-457. https://doi.org/10.3758/BF03196594

Yurchenko, A., Lopukhina, A., & Dragoy, O. (2020). Metaphor Is Between Metonymy and Homonymy: Evidence From Event-Related Potentials. Frontiers in Psychology, 11(2113). https://doi.org/10.3389/fpsyg.2020.02113

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*(4), 779-793. https://doi.org/10.1016/S0749-596X(03)00006-8

DOI: 10.1177/17470218211048386

Figure Captions

Figure 1. Ambiguity effect size (absolute differences between the activation for each of the two types of ambiguity and the activation of unambiguous words as predicted by SSD account; Armstrong & Plaut, 2016) as a function of time spent in semantic processing. The green line represents the size of the homonymy effect; the blue line represents the polysemy effect size. Area A represents an early polysemy effect, with no homonymy effect. Area B illustrates the hypothetical effect change in a prolonged processing condition, where the polysemy effect becomes attenuated, and the homonymy effect emerges.

Figure 2. Interaction between task modality, the entropy of the probability distribution, and familiarity. Faster-processing visual LDT is presented in the left panel and slower auditory task in the right panel. The left panel corresponds to the blue line in area A, and the right panel corresponds to the blue line in area B in Figure 1. Reaction latencies (negative inverse) are colour-coded in a way that the brighter shades present slower processing, and darker ones, faster processing. The green lines denote areas in which the processing latencies are the same. White areas represent values that our sample of words does not cover.

Figure 3. Interaction between prolonging manipulation, redundancy of the probability distribution, and familiarity. Faster-processing visual LDT is presented in the left panel and slower auditory task in the right panel. The left panel corresponds to the green line in area A and the right panel corresponds to the green line in area B in Figure 1. Reaction latencies (negative inverse) are colour-coded in a way that the brighter shades present slower processing, and darker ones, faster processing. The green lines denote areas in which the processing latencies are the same. White areas represent values that our sample of words does not cover.

DOI: 10.1177/17470218211048386

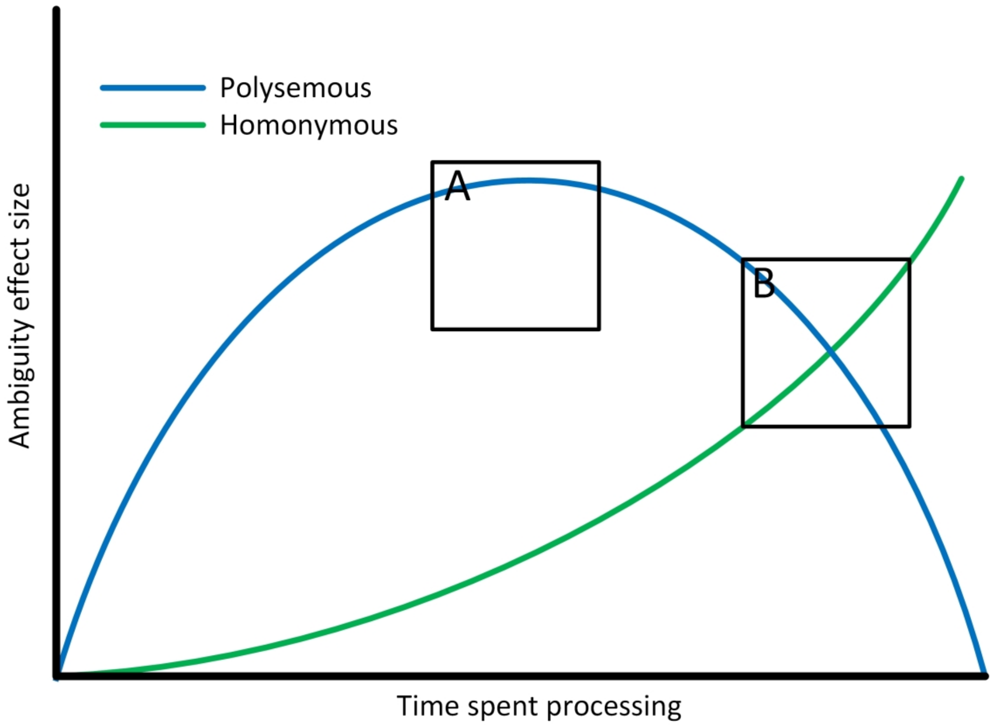

Figure 1. Ambiguity effect size (absolute differences between the activation for each of the two types of ambiguity and the activation of unambiguous words as predicted by SSD account; Armstrong & Plaut, 2016) as a function of time spent in semantic processing. The green line represents the size of the homonymy effect; the blue line represents the polysemy effect size. Area A represents an early polysemy effect, with no homonymy effect. Area B illustrates the hypothetical effect change in a prolonged processing condition, where the polysemy effect becomes attenuated, and the homonymy effect emerges.

268x197mm (300 x 300 DPI)

DOI: 10.1177/17470218211048386

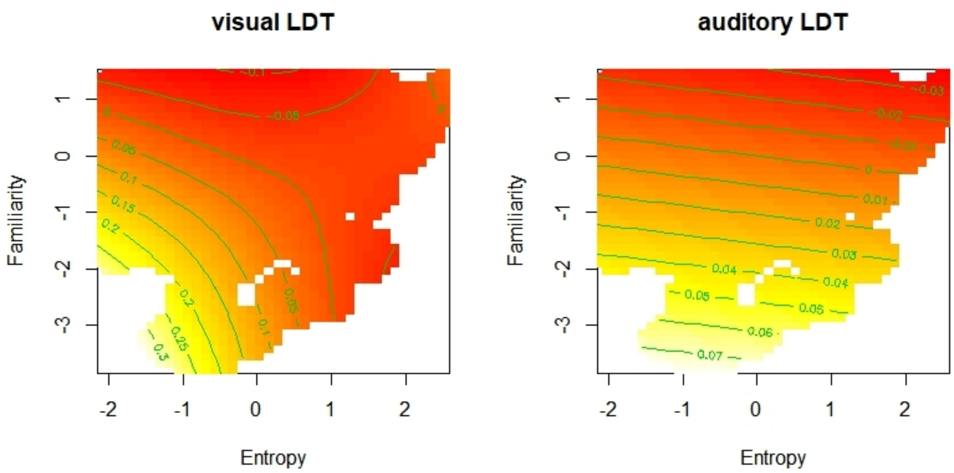

Figure 2. Interaction between task modality, the entropy of the probability distribution, and familiarity. Faster-processing visual LDT is presented in the left panel and slower auditory task in the right panel. The left panel corresponds to the blue line in area A, and the right panel corresponds to the blue line in area B in Figure 1. Reaction latencies (negative inverse) are colour-coded in a way that the brighter shades present slower processing, and darker ones, faster processing. The green lines denote areas in which the processing latencies are the same. White areas represent values that our sample of words does not cover.

212x111mm (96 x 96 DPI)

DOI: 10.1177/17470218211048386

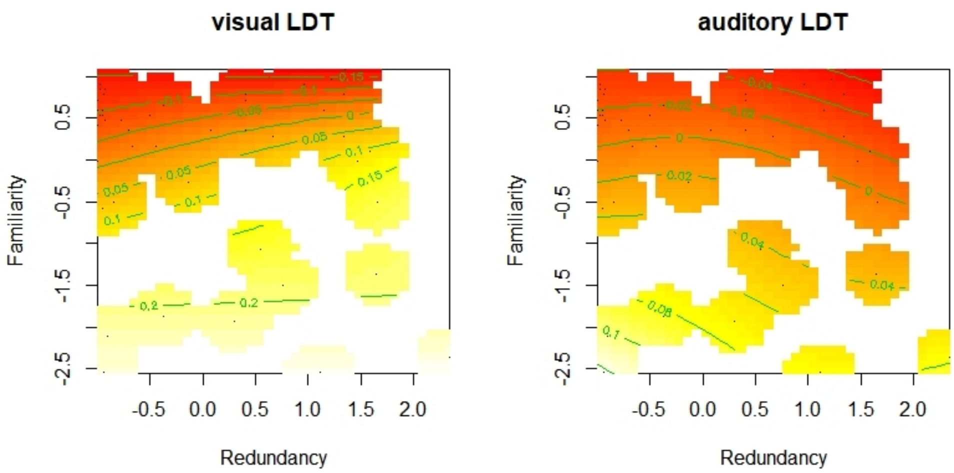

Figure 3. Interaction between prolonging manipulation, redundancy of the probability distribution, and familiarity. Faster-processing visual LDT is presented in the left panel and slower auditory task in the right panel. The left panel corresponds to the green line in area A and the right panel corresponds to the green line in area B in Figure 1. Reaction latencies (negative inverse) are colour-coded in a way that the brighter shades present slower processing, and darker ones, faster processing. The green lines denote areas in which the processing latencies are the same. White areas represent values that our sample of words does not cover.

212x111mm (96 x 96 DPI)

DOI: 10.1177/17470218211048386

Table 1

Descriptive statistics for polysemous word stimuli (taken from Serbian polysemous word norms; Filipović Đurđević & Kostić, 2017)

	M	Sd	Min	Max
Word length (in letters/phonemes)	5.04	.67	4	6
Lemma frequency[a]	204.05	214.78	10	999
Familiarity ratings[b]	6.28	.48	3.71	6.95
Word concretness ratings[c]	4.93	1.54	1.52	7
Orthographic neighbourhood size	3.63	3.86	0	18
Number of senses	11.53	4.98	4	34
Entropy	2.86	0.61	1.56	4.45
Redundancy	0.16	0.06	0.04	0.36

[a] per 2 million

[b] participants were instructed to rate the words on a seven-point Likert scale. Words rated one being least familiar and seven being the most familiar.

[c] participants were instructed to rate the words on a seven-point Likert scale. Words rated one being abstract i.e., being impossible to perceptually experience, and seven being concrete i.e., highly possible to perceptually experience.

DOI: 10.1177/17470218211048386

Table 2

Results from the generalized additive mixed model[1] fitted to response latencies for experiment 1.

Parametric coefficients:				
	Estimate	Std. Error	*t*	*Pr(>|t|)*
Intercept	-1.627	.152	-107.264	.000
(1) Trial order	.011	.004	2.765	.006
(2) Word length	.017	.004	4.021	.000
(3) (log) Word frequency	-.015	.004	-3.357	.000
(4) Orthographic neighbourhood size	.016	.004	3.781	.000
(5) Task modality (auditory LDT)	.620	.021	29.068	.000
Smooth terms:				
	edf	Ref. df	*F*	*p*
(6a) Tensor product smooth for entropy and familiarity in the baseline condition (visual LDT)	6.871	7.11	22.202	.000
(6b) Tensor product smooth for entropy and familiarity in the prolonged condition (auditory LDT)	3.000	3.00	4.214	.005
(7) By-participant factor smooth for trial order	567.745	1285.00	8.660	.000
(8) By-participant random slope for word length	50.861	142.00	0.562	.000
(9) By-item random slope for manipulation factor	248.971	285.00	7.323	.000

[1] Generalized additive mixed model summary provides information on how wiggly the regression line/surface is. Significance only states whether regression line/surface is different from zero at any point. Visualisation is obligatory for interpretation of smooth (non-linear) terms, since they are not represented by coefficients (van Rij, Vaci, et al., 2020).

DOI: 10.1177/17470218211048386

Table 3

Descriptive results of the norming study

	M	SD	Min	Max
Word length (in letters/phonemes)	4.25	1.30	3	9
Lemma frequency[a]	91.03	129.12	1	509
Familiarity ratings[b]	6.35	.56	4.93	6.97
Word concretness ratings[c]	5.73	.80	3.80	6.96
Orthographic neighbourhood size	0	18	3.63	3.86
Number of meanings	2.63	.80	2	5
Entropy	.35	.09	.20	.60
Redundancy	.10	.10	.00	.35

[a] per 2 million

[b] participants were instructed to rate the words on a seven-point Likert scale. Words rated one being least familiar and seven being the most familiar.

[c] participants were instructed to rate the words on a seven-point Likert scale. Words rated one being abstract i.e., being impossible to perceptually experience, and seven being concrete i.e., highly possible to perceptually experience.

DOI: 10.1177/17470218211048386

Table 4

Results from the generalized additive mixed model fitted to response latencies for experiment 2

Parametric coefficients:				
	Estimate	Std. Error	*t*	*Pr(>|t|)*
Intercept	-1.586	.018	-89.598	.000
(1) Trial order	.008	.004	2.259	.024
(2) Word length	.038	.009	4.378	.000
(3) Task modality (auditory LDT)	.499	.025	19.816	.000
Smooth terms:				
	edf	Ref. df	*F*	*p*
(4a) Tensor product smooth for redundancy and familiarity in the baseline condition (visual LDT)	5.830	8.90	19.738	.000
(4b) Tensor product smooth for redundancy and familiarity in the prolonged condition (auditory LDT)	3.951	4.01	3.113	.014
(5) By-participant factor smooth for trial order	309.193	1249.00	2.424	.000
(6) By-item random slope for manipulation factor	71.109	81.00	10.136	.000

DOI: 10.1177/17470218211048386