Review of PhD dissertation of *Dmytro Khahzhyn*

Dr. Emmanuel Keuleers
Department of Cognitive Science and Artificial Intelligence
Tilburg University

General Notes

- The dissertation consists of an introduction, a chapter on Semantic Processing and Semantic Priming, a chapter on Working Memory, three empirical studies (Polish Word Associations, the impact of spatial and verbal WM load on relatedness judgements, and versions of relatedness judgements of English word pairs by native and non-native speakers), a general discussion, and a conclusion.
- Taken together, the dissertation presents a coherent and impressive body of work, reflecting the candidate's ability to process and build-on existing scientific research, to perform complex experimental studies, to analyze and interpret the results of these studies, and to relate them again to existing scientific work.
- My assessment of this dissertation consists of a series of notes, remarks, and questions to the candidate. In these remarks, I address the candidate directly. Although there are minor issues with formatting here and there, I do not address these in this review.

Introduction

- The introduction is overall very well structured and gives an excellent overview to the field to semantic representations and semantic priming. A drawback of is the limited focus (isolated word processing in psycholinguistics), but it fits the context of the thesis.
- The introduction focuses on psychological theories of meaning, but pays far less attention, to relevant linguistic theories. For instance, references to the usage-based framework are largely absent. Do you agree that in many psychological studies, linguistically and philosophically important terminological distinctions do not receive enough attention, and, if you do, what do you think the consequences of that may be?
- In reading the introduction, I felt that discussions of scientific positions could be accompanied by a reference more often. While some facts may colloquially presented as trivial, that does not make them scientifically trivial. For instance, on p. 20: "Despite the fact that word meanings are almost instantaneously deciphered by speakers of a language". I understand what you mean by this, but at the same time, I know that the time course of semantic processing is a contentious issue and what "instantaneously" means in a colloquial sense obscures quite some relevant distinctions.
- On p. 17, you write that "associative relations [...] are based on the co-occurrences of words in a language [... while ...] semantic relations [...] are based on meaning overlap". This could have used some nuance, in the sense that co-occurence could be seen as an operational definition of meaning overlap. This is a point that I will come back to frequently in my assessment.

- Do you agree that the distinction between associative and semantic priming in psychology is based on a superficial understanding of "semantic relatedness". For instance, dog and bark would not be considered to have a strong semantic relationship, but would be considered to have an associative. This is wrong, both in formal and usage-based analyses of meaning (see for instance, it is used as an example of a semantic relationship on p.21 when discussing feature-based models).
- Section 1.2 discusses word meaning from a very broad perspective, and in doing so makes some points which are either contentious or could be tied to existing linguistic concepts. For instance, the claim that "Nothing in the word itself suggests its meaning and there is no one-to-one correspondence between the form of the word and the concept that it represents." could be tied to Saussure's notion of arbitrariness of the sign.
- On p. 22, you explain that "One of the advantages of feature-based models is that they can readily account for non-linguistic input from different modalities, such as vision, hearing, and touch". Explicitly saying "non-linguistic" makes the statement contentious, especially in the embodiment-perspective, which is discussed immediately after, which would construe what is linguistic much more broadly.
- I found that, in several places, the presentation was lacking some mathematical sophistication. For instance, on p. 28 you write that "The LSA model further employed log transformation of frequencies and a dimensionality reduction technique to generate semantic representations". While the function of dimensionality reduction (SVD) is clear, I don't think it's accurate that the other major improvement is log transformation. The tranformation is not just log frequency, but log(context frequency)/(word entropy) which relates the co-occurrence to the background rate of co-occurrence for that word.

Chapter 2

- The literature study on working memory follows the extensive literature on the subject and relates working memory to language processing. Like the introduction, I think this chapter is narrowly constrained to the field of psychology, but this fits the scope of the thesis.
- On p. 74, you write that "For example, De Deyne (2019) showed that English word association norms are better predictors of human performance on semantic tasks than word frequencies.". Since a word frequency is not a semantic measure, how can it be compared to a word association measure? Or are you referring to co-occurence frequency?

Chapter 3

- In chapter 3, you conclude that there is a distinction between associative and semantically related words, based on the finding that, "for the majority of cues, semantically related target words from Rataj et al. (2023) never appeared among association responses." If I understand this correctly, the semantically related words you are referring to are the strongly and weakly associated prime from Rataj.

- This reasoning crucially depends on the probability that your semantically related words will be generated in an association study of the kind you performed. I would want you to consider how certain that is, especially since you only consider two words to be semantically related per cue word.
- Let's consider a very conservative thought experiment: for each cue word there exist 32 associatively related responses, three of which are selected by a stochastic random sampling process without replacement, with some responses being more likely than others. The most likely response has a probability of 0.20, the 2nd and 3rd have a probability of 0.10, and so on. Since responses are mutually exclusive, the probability of a response occurring in a 3-association trial, is its response probability * 3. The table below illustrates this.

response rank	number of responses	response probability	total probability mass for rank	word set	Approximate probability of a *specific* word with this rank occuring in a single 3-association trial
1	1	0.2	0.20	[1]	0.60
2	2	0.1	0.20	[2:3]	0.30
3	4	0.05	0.20	[4:7]	0.15
4	8	0.025	0.20	[8:15]	0.075
5	16	0.0125	0.20	[16:31]	0.0375

- Given such a situation, how likely would you be to find the strong and weak semantic related word in an association task? What if there were more than 32 possible associative responses?
- In the context of chapter 3, did you consider the relationship between specific models of constructing word embeddings (e.g., CBOW) and the degree to which similarity between the resulting vectors could be associative in nature? Are there indications that the vector space models used by Rataj would be particularly (un)suited at encoding associative relations? Would there be grounds to say that the *semantically* related stimuli in Rataj were (also) associatively related in the first place?
- One could say that *associative* is a task-related term (words humans produce when given a cue word), whereas *semantic* is a linguistic concept referring to a wide variety of relationships based on the meaning of a word. Do you agree with this *distinction* and could you elaborate on why it may be relevant for interpreting empirical results?
- Continuing on the previous remark: models of semantic similarity allow you to establish the semantic similarity between any pair of words, whereas methods to establish associative similarity depend on a task generating process. If not semantic, what type of relation would associations be based on?
- You explain that a word association task carries the risk that participants associate their subsequent responses to previous responses instead of to the cue word (so-called response chaining). You use a statistical test to dispel this concern, showing that second and third responses are more related to the cue word than to the previous responses.

But could it be the case that vectors for response words simply tend to be more distant in the vector space for other reasons? There are good reasons to believe this may be the case: the less frequent a word is, the less information is encoded by its vector, and, therefore, the more distant it will be to other vectors on average. How would you establish whether this is the case, and, if so, how would correct your test?

- In this context, is the fact that you don't find a word with a *weak* semantic link any indication that associative and semantic relations are different?
- Do you think it would have been worthwhile to test where the similarity between cue and each of the associations rank *with respect to* the similarity between cue and strong and weak semantic targets?
- If semantic and association similarity were truly independent, would you still predict semantic similarity between cue and associative responses to decline with response order?
- At the end of chapter 3, you present two things which are seemingly in contradiction: On the one hand, you state that the "stimuli [from Rataj] can be used to specifically investigate semantic relations". On the other hand, you state that "measures of associative relatedness [...] correlate with the measure of semantic similarity". Do you think this requires some nuance?

Chapter 4

- Chapter 4 describes a series of well though-out experiments that address the question how working memory load (low vs high) and type (spatial vs verbal) affects semantic relatedness judgements for word pairs with different degrees of semantic relation.
- The hypothesis is that both high (vs low) working memory load and verbal (vs spatial) memory load would make semantic relatedness judgements slower and less accurate.
- Although you took care to make sure the spatial memory task would not involve verbal memorization of the location by presenting 8 positions on a circle, rather than 4 orientations, wouldn't it be possible for participants to try to remember these positions as approximate clock readings (e.g. 11) or bearings (SSW). If this were the case, how do you think it could have influenced your findings?
- Do you think a true "verbal" version of the verbal working memory task have had a different effect (e.g., doing the same/different task with verbal stimuli presented through headphones)?
- I was surprised by some aspects of the statistical analysis:
    - "Generalized linear models were used because they do not require the assumption that RTs are normally distributed". This seems to assume that raw data need to be normally distributed in the first place for the general linear model. I was taught that this is incorrect and that only the residuals need to be normally distributed. What is your view on this?
    - Do you agree that generalized linear models are not a panacea that will allow us to handle any kind of aberrant data and that the characteristics of the data always need to be carefully considered?

    - You write that "because models with random slopes failed to converge, the final model included only random intercepts for subjects and items.". In my experience, often, a model failing to converge, is the result of a mismatch between the design of the experiment and the model specification. What would you consider the "fully specified" model, given the design of your experiment?
- The second and third experiment show that the effects of degree of relatedness occur independently of the working memory task, and whether a delay is present between presentation of the target word and the judgement. They also establish *baseline* levels for speed and accuracy in the semantic relatedness judgement task. Do you think comparing these baseline levels directly to the results obtained in experiment 1 would give additional insights?

Chapter 5

- Chapter 5 investigates semantic relatedness judgements by native and non-native speakers. It also specifically compares how forward, backward, and symmetric associates compare in terms of speed and accuracy of semantic relatedness judgements. The results, are broadly in line with the results from Chapter 4.
- The lack of attention to the, at least epistemological, differences between association and semantic similarity in this chapter is somewhat surprising, given the insistence in Chapter 3, that semantic similarity and word association are distinct and that stimuli having no associative relationship is crucial to making conclusions about semantic processing. It appears here that a much more nuanced way of phrasing the relationship between word association and semantic similarity is required, or it wouldn't be possible to interpret the results of chapter 5 of being relevant to semantic processing at all. What is your view on this?
- On p. 146 you write that "Statistical analyses did not reveal significant differences in the facilitation effects between groups although the distributional analysis showed a tendency towards larger facilitation effects in non-native as compared to native speakers, especially for faster responses. This pattern of results may indicate a trade-off between the speed of relatedness judgements and the spreading activation with words presented in the non-native language." How would you interpret this result in the context of the studies done by, for instance, Kuperman and van Dyke (2013) or Cop et al. (2015)
    - Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review, 22*(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2
    - Kuperman, V., & Van Dyke, J. A. (2013). Reassessing Word Frequency as a Determinant of Word Recognition for Skilled and Unskilled Readers. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/a0030859

Chapter 6

- Chapter 6 summarizes the results of the thesis in an accessible manner.

- Again, the author addresses the difference between associative and semantic relations and how views within the literature have evolved. However, the author does not seem to commit to any of these interpretations. Given the experience that the candidate undoubtedly accumulated during the writing of this dissertation and the empirical results obtained in the thesis, it would be evident to ask the author to formulate and defend their position.

Overall Assessment

Overall, despite critical comments, I give the dissertation a clear positive assessment. My evaluation leads me to believe that the candidate can proceed to the final stages of earning a PhD title.

18/09/2023

Emmanuel Keuleers