Computational Psycholinguistics Assignment 1: Word2Vec and Semantic Priming

Name student 1: Rong Ding
Student Number: s1067006

Name student 2: Junyuan Zhao
Student Number: s1058733

In this assignment, we investigated how well cosine values generated by Word2Vec models simulated 1) human semantic categorisation of word pairs as reflected in Nelson et al. (2004), and 2) predicted semantic priming effects in Hutchinson et al., (2013). We also trained models with different parameters to investigate the effects of model parameters and their interactions on model performance.

1. Model Training

1.1 Data Cleaning

To begin with, we first trained a *word2vec* model with default parameters on the ENCOW corpus (Schäfer, 2015). Prior to training, punctuations (e.g, “!”, “?”) were identified and removed from the corpus with a regular expression. This way, only symbols that occurred within words remained in the corpus. All models that we report below were trained on the modified corpus, which was saved as *train_en_depunc.txt.*

1.2 Parameter selection

We decided to manipulate the following parameters: *size* and *window*. For the *size* parameter, the values 50, 100 and 150 were used. For the *window* parameter, the values 3, 5 and 7 were used. Altogether, we had 3 × 3 = 9 models in total.

2. Data analysis

2.1 Materials

Datasets used in the current study were obtained from the study by Hutchison et al. (2013).

2.2 Data preprocessing

Following training of each model, the original experimental datasets of the naming and lexical decision tasks were each loaded into *word2vec_training.ipynb* as a Dataframe using the *pandas* package. A cosine similarity value was then computed for each pair of prime and target and added into the Dataframe as a column named *cosine*. For trials whose prime or target was absent in the model vocabulary, the cosine similarity was marked as an *NaN* (i.e., missing value), and those trials were subsequently removed from the Dataframe. The Dataframe was then extracted and saved as an intermediate csv-file, and imported into *Data Preprocessing.ipynb* for preprocessing.

During preprocessing, the previously produced metafile was again loaded as a *pandas* Dataframe named *data_naming* (or alternatively, *data_lexdec,* hereinafter, for lexical decision data). To preserve the match between conditions in each experimental dataset, all of the trials with a particular target were removed from *data_naming* if any trial including the target had been removed due to missing similarity scores previously (see cells under *Step 1*). At this point, a separate datafile was saved from the current Dataframe for statistical analysis of cosine similarities (*cosine*). We further computed differences in mean response times (*meanRTs*) and cosine similarities between conditions to help establish statistical relationships between semantic priming effects and prime-target cosine similarity. Specifically, we first computed the priming effect of meanRTs which was quantified as the difference between *strong* versus *unrel_strong* (or *weak* versus *unrel_weak*) conditions for each target with each specific *isi*. We also computed the difference between cosine similarities of *strong* versus *unrel_strong*, as well as *weak* versus *unrel_weak* conditions per target. Overall, priming effects of meanRTs and cosine differences can be represented with the following equations (the direction of subtraction for priming effects followed Hutchison et al., 2013):

Priming effect(meanRT) = meanRT_unrel_strong(/unrel_weak) - meanRT_strong(/weak)
Difference(cosine) = cosine_strong(/weak) - cosine_unrel_strong(/unrel_weak)

Priming effects and cosine similarity differences were computed on the basis of *data_naming*. A new Dataframe named *data_primingeffect* was created to save the computed results, wherein priming effects and cosine differences were stored respectively as *prim_diff_RT* and *diff_cosine* (see cells under *Step 2*). Variables such as isi, condition, prime of the strong/weak condition, prime of the unrel_strong/unrel_weak condition and target were also saved along with the results in each row for subsequent data analysis. We included in **Table 1** below a description of the name and meaning of variables in each row of *data_primingeffect*. In the final step, to exclude the impact of potentially overly influential outliers, we removed data points in *data_primingeffect* based on a criterion of ±2 SD (see cells under *Step 3*). Specifically, we first classified the dataset into 4 different subsets in accordance with the 4 isi-condition combinations, and then z-transformed the values of *prim_diff_RT* and *diff_cosine* in each subset. Data points were excluded if the z-score of *prim_diff_RT* or *diff_cosine* exceeded 2. Remaining data in the *data_primingeffect* (i.e., an average of 6006 data points in total per model for the lexical decision task, and 4972 for the naming task; see **Appendix Table 1** for detail) were then extracted and saved as a csv-file for statistical analyses.

Table 1: Names and description of variables (columns) in *data_primingeffect*

Variable (Column name)	Description
isi	inter-stimulus interval of the contrasted trials
prime	prime word of the contrasted trial of the *strong*/*weak* condition
prime_un	prime word of the contrasted trial of the *unrel_strong*/*unrel_weak* condition
target	target word of the contrasted trials
condition	the strong/weak semantic relationship (condition) upon which the computed priming effect is based; coded as *strong* or *weak*
prim_diff_RT	difference (i.e., priming effect) between meanRTs of strong/weak and corresponding unrelated conditions; quantification criterion has been shown above
diff_cosine	difference between cosine similarities of strong/weak and corresponding unrelated conditions; quantification criterion has been shown above

All the aforementioned preprocessing steps were accomplished using *Anaconda 4.10.0*, *Python 3.7.4*, and the *Pandas* package (version 1.2.3). Both ipynb-files have been made available along with the current document.

2.3 Data Analyses

Data analyses were conducted on the lexical naming and decision datasets separately, yet they followed identical analysis pipelines which we delineate in the following paragraphs of this section.

First, to check whether the cosine similarity values from each model we trained matched categories of semantic relatedness in the task stimuli, we conducted a one-way analysis of variance (ANOVA) on the cosine similarity data across conditions. Cosine similarities between primes and targets were the dependent variable. The independent variable was *condition*. Contrasts were made between *strong*, *weak*, and *unrelated* (i.e., *unrel_strong* and *unrel_weak*) conditions. Note that only the data with a 50-ms *isi* were used in the ANOVA, as prime-target pairs (and therefore cosine similarity values) were identical in trials with 50- and 1050-ms *isi*'s. Furthermore, we extracted F-values of the ANOVAs as an index of how accurately a model differentiated between the three semantic-relatedness categories. We compared the F-values to investigate the effects and interactions of different model parameters.

Next, linear mixed-effects regression (LMER; Baayen et al., 2008) modelling was performed to evaluate how well the size of the priming effect was predicted by cosine similarities. For each of the nine trained models, we included *prim_diff_RT* (i.e., priming effect in meanRTs) as the dependent variable, and *diff_cosine* (i.e., difference in cosine similarities) as the predictor variable of interest. We also put *isi* among fixed effects as a control variable to account for the variance brought about by variations in response patterns due to task demand. *Target* and *prime* were included in the random structure as intercepts to account for variance which had been introduced by particular word stimuli. Estimated p-values for fixed effects were computed using Satterthwaite's approximation for degrees of freedom implemented in the *lmerTest* package (Kuznetsova et al., 2017). Assumptions of LMER modelling were checked by visual inspection on residual distribution plots. The model structure can be represented with the following equation:

$$prim\_diff\_RT \sim diff\_cosine + isi + (1 \mid target) + (1 \mid prime)$$

After model implementation, we estimated marginal R^2 for the coefficient of *diff_cosine*, that is the variance of priming effects explained by cosine similarity differences solely. Similar to the F-values from ANOVAs, we see the R^2-values as a representation of how accurately cosine similarity differences predict or explain variations in the semantic priming effect size. Therefore, we extracted the R^2-values of *diff_cosine* from the nine LMER models performed on the datasets of the nine trained *word2vec* models. We then plotted the variation of R^2-values against *window* and *size* and, together with F-values, we analysed the influence of the two factors on *Word2Vec* model performance in predicting human semantic associative capabilities.

All the analyses in the current study, including ANOVA and mixed models, were performed under RStudio (version 1.4.1106). ANOVA and multiple comparisons were conducted using the R package *ggpubr* (version 0.4.0) and *multcomp* (version 1.4.16). LMER modelling was implemented using the package lme4 (Bates et al., 2014; version 1.1.26). Residual distribution plots used for assumption check were created using the package *effects* (version 4.2.0) and *sjPlot* (version 2.8.7). R^2 estimation was conducted using the package *piecewiseSEM* (Lefcheck, 2016; version 2.1.2). Codes have been made available along with the current document.

2.4 Predictions

With respect to the ANOVA analyses on cosine similarities, we expected to see significant differences between the three conditions of semantic closeness, which should serve as an indication that cosine similarities reliably predict human semantic categorisation. With regard to the mixed regression between priming effects and cosine similarity differences, we also expected that differences in cosine similarities would significantly contribute to the variance in priming effects; concretely, we expected cosine similarity differences to be positively correlated with priming effect sizes.

3. Results

3.1 Word Naming Data

3.1.1 Experimental conditions

For how well the distribution of cosine similarities matched experimental conditions, we only report here the results from a default *Word2Vec* model (*window* = 5, *size* = 100). We delay a discussion of the effects of ISI and parameter selection until section 3.1.2 and 3.1.3, respectively.

A one-way ANOVA yielded a significant main effect of condition, $F(2,5469) = 2228$, $p < 0.001$. A follow-up post-hoc test with Tukey's HSD method yielded significant differences among conditions. Cosine similarities of prime-target pairs under the "strong" condition were the highest ($M = 0.49$, $SD = 0.22$) as compared to the "weak" condition ($t = 16.19$, $p < 0.0001$) and the "unrel" condition ($t = 62.22$, $p < 0.0001$). Cosine similarities of prime-target pairs under the "weak" condition were intermediate ($M = 0.36$, $SD = 0.22$), larger than the "unrel" condition ($t = 43.53$, $p < 0.0001$). Prime-target pairs under the "unrel" condition had the lowest cosine similarities ($M = 0.08$, $SD = 0.16$). This means that the distribution of cosine similarities well matched the conditions in the actual experiment, and thus provided convincing evidence that cosine similarities yielded by our trained *Word2Vec* model well simulated human semantic categorisation of word pairs.

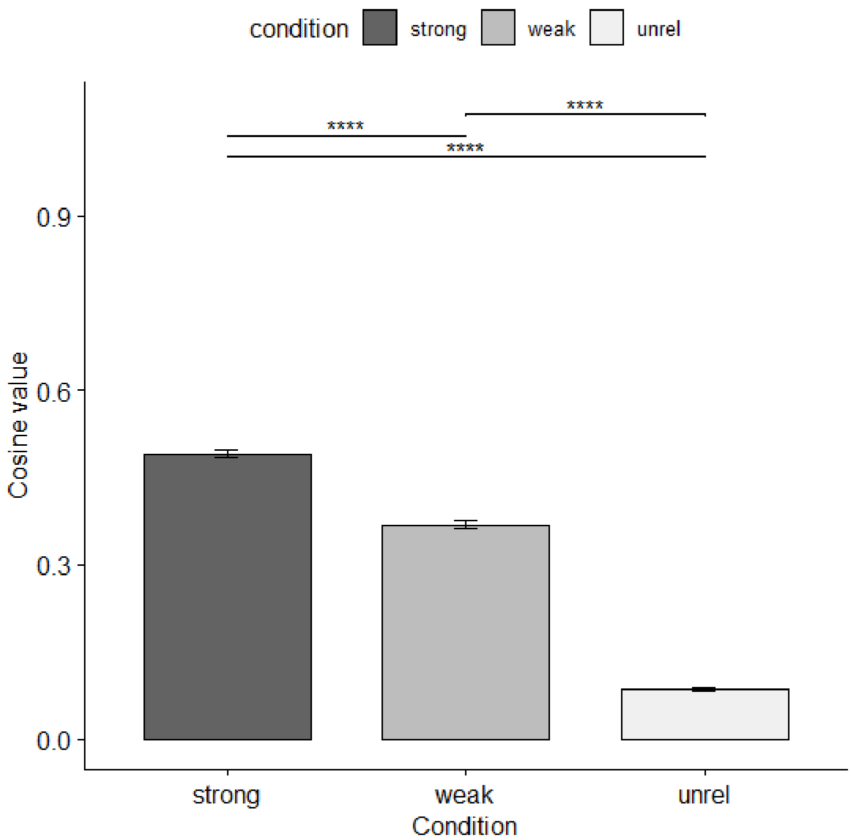

Figure 1: Mean cosine similarities by condition (ISI = 50, word naming data)

(significance level: *: p < 0.05, **: p < 0.01, ***: p < 0.001, ****: p < 0.0001)

3.1.2 Priming Effects

Similar to 3.1.1, we only report here the results of a default model. We delay a discussion of the effect of parameter selection until section 3.1.3.

As for fixed effects, the term *diff_cosine* showed a positive correlation with *prim_diff_RT*, which reached significance under a t-test using Satterthwaite's estimation method ($\beta = 7.47$, $t(3550.4) = 2.91$, $p = 0.004 < 0.01$). The term *isi* did not reach significance under a subsequent t-test. Random slope terms in our model did not explain much of the residuals (for details see Appendix Table 2).

The positive correlation between *diff_cosine* and *prim_diff_RT*, at the surface, meant that the larger cosine differences between prime-target pairs (e.g., between “ability-capability” and “ability-fuel”, **not** between prime and target themselves), the larger priming effects. Since cosine similarity differences are calculated by subtracting cosine similarities of strong/weak conditions from that of unrelated conditions (i.e., a baseline), larger differences in cosine similarity indicate closer semantic relationship **within** the prime-target pairs of interest. To make it more concrete, our modelling results indicate that a prime that is semantically more similar to the target primes the target more strongly. Regarding *isi*, the assumption of Hutchison et al. (2013) that different inter-stimulus intervals might entail different cognitive processes was not borne out in the naming data, since including *isi* as a fixed effect term did not explain much of the variance in data. This indicates that the word naming (actually, word reading) task is not sensitive to differences in inter-stimulus interval.

3.1.3 Effect of Model Parameters

Plots of F-values of all nine ANOVA models and marginal R^2 from all nine mixed effect models on the word naming data have been displayed below (see Fig.2 and Fig.3, respectively). Both hidden layer size (as determined by the *size* parameter) and number of neighboring words (as determined by the *window* parameter) modulated model fit. Given the parameter space that we had, we observed that *window* and *size* both had a constantly positive effect on F-values. However, when it comes to R2, *window* and *size* interacted in a more complicated way.

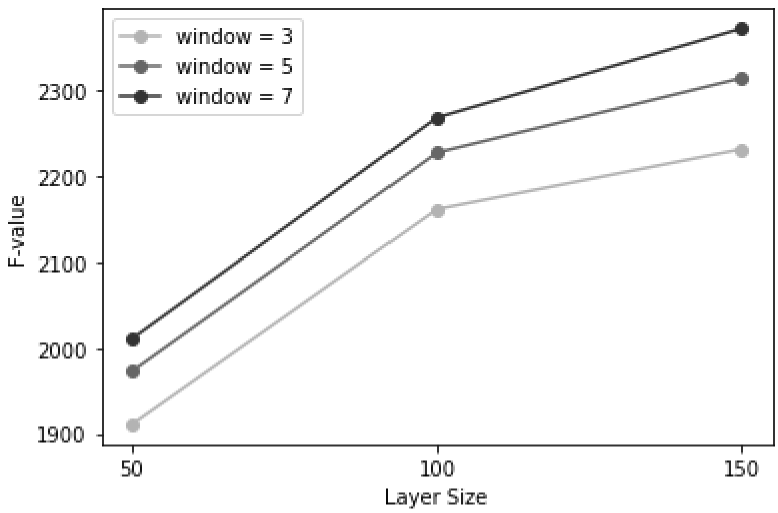

Figure 2: F-values from ANOVAs (*isi* = 50, naming data)

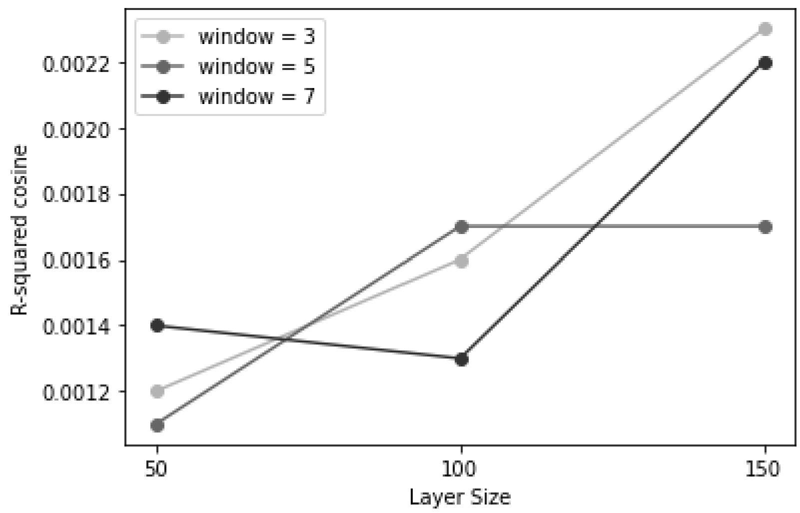

Figure 3: Marginal R^2 of LME models (naming data)

As is evident in Figure 2, as *window* increased, F-values of ANOVA increased. Increasing F-values of ANOVA performed on naming data with an ISI of 50ms indicate that the distribution of cosine similarities generated with larger hidden layer sizes matched experimental conditions better. However, R^2 values from LME models did not show a monotonic increase as F-values did. To begin with, we observed a global trend of increase in R^2 with growing layer size, irrespective of *window* size. This may suggest that increasing dimensions of semantic features in the model improved model performance, although the causes remain elusive. Assuming that a larger hidden layer size leads to better simulation of real-world semantic distributions given a certain *window*, we infer that within the current parameter space, a *window* of 3 may better reflect human cognitive capacities (e.g., working memory), at least those subserving word naming. However, the reason behind the interaction between *size* and *window,* which led to different concaves is still unclear to us. Further exploring the computational details underlying *Word2Vec* models is beyond the scope of this assignment.

3.2 Lexical Decision Data

3.2.1 Experimental conditions

Fig.4 reports the results of comparisons among cosine similarities of strong, weak, and unrelated conditions in the lexical decision task we extracted from the default model. Similar to our findings on the naming dataset, we found a significant effect of condition ($F(2, 6553) = 2614$, $p<0.0001$). Follow-up tests using Tukey's post hoc test revealed that cosine similarities of prime-target pairs in the strong condition ($M = 0.4877$, $SD = 0.2261$) were larger than those in the weak condition ($M = 0.3660$, $SD = 0.2243$) ($t = 17.69$, $p<0.0001$), which in turn were greater than cosine similarities between semantically unrelated primes and targets ($M = 0.0857$, $SD = 0.1639$) ($t = 47.04$, $p<0.0001$). These results confirmed that our choice of modelling parameters had effectively categorised the prime-target pair stimuli used in the decision task on the basis of their semantic relatedness. Therefore again, our results suggested that simulation performance of our trained model to meaning categorisation of words or word pairs was adequate.

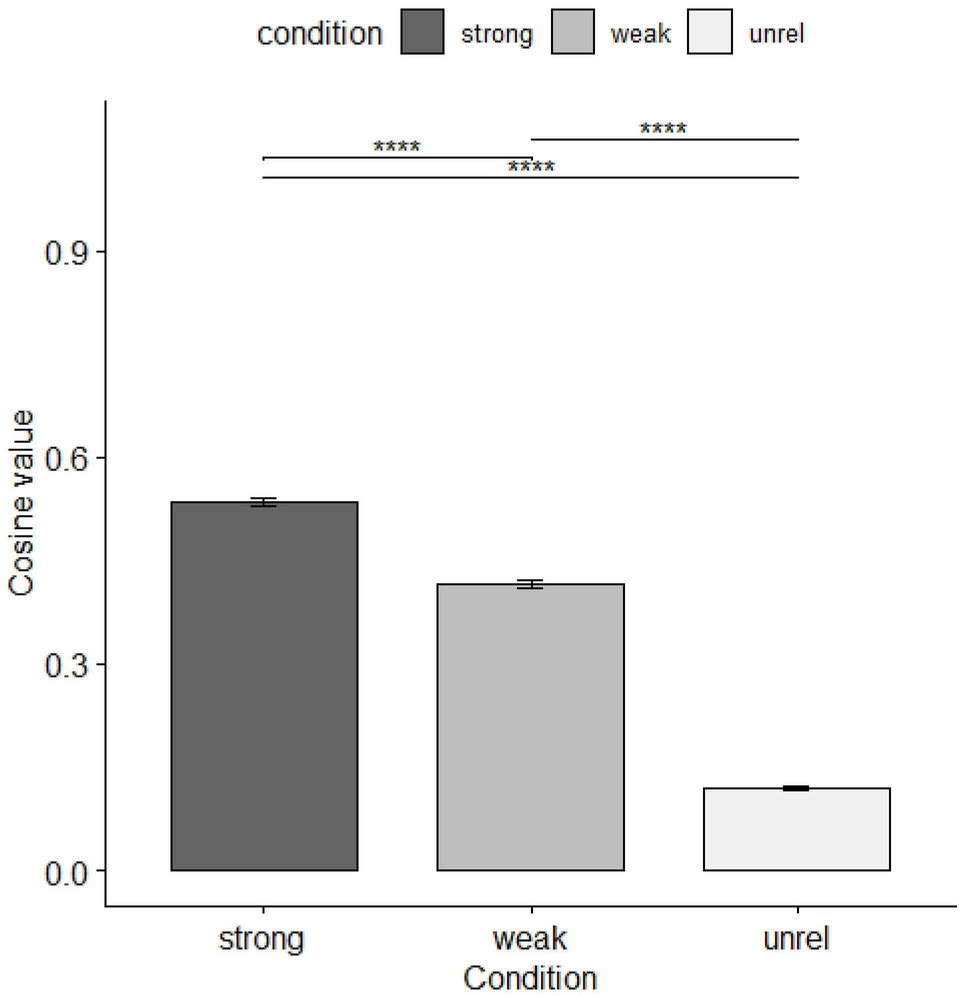

Figure 4. Mean cosine similarities by condition (ISI = 50, lexical decision data) (significance level: *: p < 0.05, **: p < 0.01, ***: p < 0.001, ****: p < 0.0001)

3.2.2 Priming Effects

Again, by carrying out LMER modelling on priming effects and cosine similarities produced by the default model, we observed a significant effect of cosine similarity difference (β=20.897, t=6.084, p<.001), that is, priming effect size increased as differences in cosine similarities between primes and targets increased. These results indicated that greater facilitative effects in meanRTs were associated with closer semantic relationships between primes and targets. Interestingly, different from what we observed in the naming dataset, *isi* showed here in the lexical decision task a significant effect on priming effect size (β=6.61, t=3.946, p<.001). Moreover, it is also important to note that the *isi*'s in all the LMER models on the decision task datasets showed a significant effect (p's<.001; see Appendix Table 2 for detail), whereas *isi* didn't show significance in any LMER model on the naming task datasets. Therefore, we see such consistent differences in the effects of *isi* as an indication that task-specific demands exerted influence on the size of semantic priming effects.

Details about the results of the LMER model described here can be found in Appendix (see Appendix Table 2) along with results of the LMER models performed on cosine similarities yielded by other trained models.

3.2.3 Effect of Model Parameters

Plots of F-values of all nine ANOVA models and marginal R^2 from all nine mixed effect models on the word naming data have been displayed below (see Fig. 5 and Fig.6 respectively). Again, both hidden layer size (i.e., the *size* parameter) and number of neighboring words (i.e., *window* parameter) modulated model fit. However, while F-values here showed patterns which were highly similar to those of the F-values on the basis of ANOVAs on the naming data, R^2 yielded patterns which were as intricate but were also different from those extracted from LMER models on the naming data.

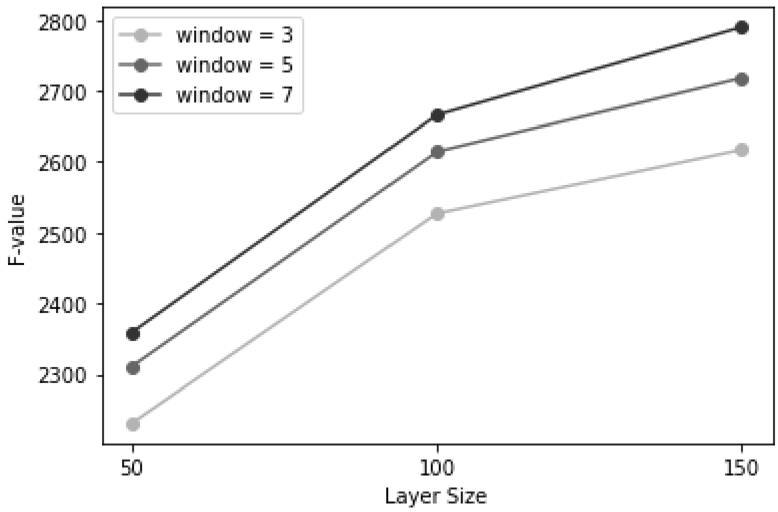

Figure 5: F-values from ANOVAs (*isi* = 50, lexical decision data)

As shown in Fig.5, positive modulation of both *window* and *size* have been observed on F-value: the F-value increased with the enlargement of the size of the window or hidden layers. This indicated that our models were better-performed in semantic association of words when the hidden layer and window sizes were larger. In contrast, patterns of interactions between *window* and *layer size* seem more complex in marginal R^2. Our results of the lexical decision task data at layers sizes of 50-100 showed highly similar patterns with results of the naming task data: R^2-values of cosine similarity differences from models with a smaller window grew with the increase of layer sizes. However, at the layer size of 150 we see contrary patterns: the marginal R^2-value of the model with a window size of 7 seemed to outgrow those of models of window sizes 3 and 5. This might instead indicate that the low performance of models with a window size of 7 at smaller layer sizes as in the data of both tasks was attributed to the lack of an adequate amount of semantic dimensions to encode word associations, which might be reasonable though, considering that F-values overall increased as the window became larger. Such results pointed to the potential limitation of smaller parameters in observing and evaluating model performance.

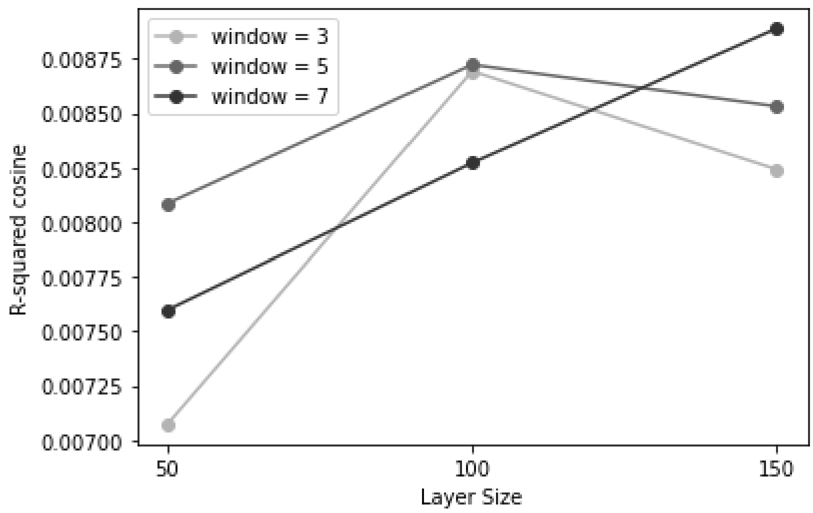

Figure 6: Marginal R^2 of LME models (lexical decision data)

Apart from task-internal observations, we also noticed a systematic difference across tasks—that is, both R^2 and F-values for naming data were overall numerically smaller than those for the lexical decision task. We attribute the difference to experimental task specificity. We noticed in Hutchison et al. (2013) that the "word naming" task is indeed "word reading", rather than a more common "picture naming". Word reading requires participants to generate articulatory responses given orthographic inputs. We thus speculate that this specific experimental task would rely more on phonological processes such as syllabification, as compared to semantic processes which are necessarily engaged in picture naming (for an overview of task demands of picture naming, see Indefrey & Levelt, 2004). Consequently, in word reading (pronunciation) tasks, semantic factors might be involved, but more as a result of passive spread activation from orthographic activation. In contrast, lexical decision tasks potentially involve more complete lexical access. An input needs to activate the corresponding orthographic nodes, which directly feed forward to semantic nodes, to a threshold to be recognized. Semantic factors are thus more activated in lexical decision tasks, leading to more pronounced semantic effects. Since distributional word vectors (and cosine similarities computed based on them) mainly carry semantic information, it is reasonable that we observe less variance, as indicated by smaller R^2, in the naming data.

4. Conclusion

In this assignment, we ran Word2Vec models on the ENCOW corpus, and modelled experimental semantic priming data (Hutchison et al., 2013) with word cosine similarity values generated with Word2Vec models. We manipulated model parameters *window* and *size* while keeping other parameters as default. Each model's goodness-of-fit was evaluated in terms of F-values from ANOVA, and R^2 from LME models.

As predicted, the distribution of cosine similarities matched the three experimental conditions in Hutchison et al. (2013). We also found a positive correlation between the difference in cosine similarities and semantic priming effects. These basic findings confirmed that distributional word vectors well captured semantic properties of words.

We observed different model coefficients between two experimental datasets (i.e., word naming and lexical decision), with R^2 (which reflects the amount of variance explained by cosine values) higher in general for lexical decision data. Moreover, we did not observe a significant fixed effect of *isi* in the word naming data, as compared to in the lexical decision data. We speculate that these findings reflected task-specific cognitive underpinnings. By manipulating model parameters, though, we obtained unpredicted yet intriguing results suggesting that *window* interacted with *size* in a more complicated way, which was hard to explain given the scope of our experiment. Our hypothesis given current data could only be that a *window* of 3 ensembles cognition to a greater extent.

We acknowledge that there are several limitations in our assignment. First, we did not run any models with larger *window* and/or *size* as control, and thus failed to identify whether our current findings were specific to *Word2Vec* models of small parameters. Second, in LMER modelling we did not further control for each word several key confound variables such as word frequency, word length, semantic diversity, or interference from orthographic or phonological neighbors (which, though, has been commonly practiced in research using LMER model; e.g., Rotaru, Vigliocco, & Frank, 2018). Third, we did not directly compare cosine values in the naming and lexical decision tasks, and therefore could not answer whether the differing patterns observed in naming and lexical decision data were due to differing patterns of cosine similarity distribution.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R Package Version, 1(7), 1–23.

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C.-S., Yap, M. J., Bengson, J. J., Niemeyer, D., & Buchanan, E. (2013). The semantic priming project. Behavior Research Methods, 45(4), 1099–1114. https://doi.org/10.3758/s13428-012-0304-z

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. Cognition, 92(1), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13). https://doi.org/10.18637/jss.v082.i13

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods in Ecology and Evolution, 7(5), 573–579. https://doi.org/10.1111/2041-210X.12512

Rotaru, A. S., Vigliocco, G., & Frank, S. L. (2018). Modeling the Structure and Dynamics of Semantic Processing. *Cognitive Science*, *42*(8), 2890–2917. https://doi.org/10.1111/cogs.12690

Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/3836

Appendix

Table 1: Total number of trials included in statistical analyses of each model

Model name	Data modelled	Parameter: window	Parameter: size	N_trials
model_naming_window3size50	data_naming.csv	3	50	4988
model_naming_window3size100	data_naming.csv	3	100	4964
model_naming_window3size150	data_naming.csv	3	150	4958
model_naming_window5size50	data_naming.csv	5	50	4980
model_naming_window5size100	data_naming.csv	5	100	4978
model_naming_window5size150	data_naming.csv	5	150	4954
model_naming_window7size50	data_naming.csv	7	50	4974
model_naming_window7size100	data_naming.csv	7	100	4972
model_naming_window7size150	data_naming.csv	7	150	4979
model_lexdec_window3size50	data_lexdec.csv	3	50	6021
model_lexdec_window3size100	data_lexdec.csv	3	100	5994
model_lexdec_window3size150	data_lexdec.csv	3	150	5987
model_lexdec_window5size50	data_lexdec.csv	5	50	6018
model_lexdec_window5size100	data_lexdec.csv	5	100	6011
model_lexdec_window5size150	data_lexdec.csv	5	150	5980
model_lexdec_window7size50	data_lexdec.csv	7	50	6016
model_lexdec_window7size100	data_lexdec.csv	7	100	6012
model_lexdec_window7size150	data_lexdec.csv	7	150	6013

Table 2: Full results of LMER modelling implemented on 9 trained models

Lexical Naming data					
Default model (win=5, size=100)					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-4.3679	1.2795	3672.8073	-3.414	<.001
diff_cosine	7.4703	2.5713	3550.4255	2.905	<.01
isi	-0.7556	1.2860	3735.6179	-0.588	0.557
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	0	0		
target	(Intercept)	7.318	2.705		
Residual		2065.386	45.447		
win=3, size=50					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-4.788	1.241	3729.573	-3.859	<.001
diff_cosine	5.344	2.297	3670.797	2.326	0.020
isi	-1.116	1.287	3743.101	-0.867	0.386
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	0	0		
target	(Intercept)	7.318	2.705		
Residual		2065.386	45.447		
win=3, size=100					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-3.9813	1.2746	3657.2917	-3.123	<.01

diff_cosine	9.1684	2.7919	3529.5647	3.284	<.01
isi	-0.9259	1.2902	3713.9559	-0.718	0.473
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	0	0		
target	(Intercept)	10.71	3.273		
Residual		2062.57	45.416		
win=5, size=50					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-4.9192	1.2548	3687.0372	-3.920	<.001
diff_cosine	5.2590	2.2862	3553.5242	2.300	0.0215
isi	-0.9562	1.2861	3731.1565	-0.743	0.4572
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	0	0		
target	(Intercept)	4.114	2.028		
Residual		2058.891	45.375		
win=5, size=150					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-4.3212	1.2873	3659.5709	-3.357	<.001
diff_cosine	7.8638	2.7736	3519.2112	2.835	<.01
isi	-0.8589	1.2895	3714.5788	-0.666	0.505
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	0	0		
target	(Intercept)	8.745	2.957		
Residual		2058.956	45.376		
win=7, size=50					
Predictor (Target/Control	**Estimate**	**S.E.**	**df**	**t**	**p**

variable)					
(Intercept)	-4.606	1.264	3712.586	-3.645	<.001
diff_cosine	5.738	2.287	3647.566	2.509	0.012
isi	-1.167	1.285	3726.724	-0.908	0.364
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	0	0		
target	(Intercept)	9.587	3.096		
Residual		2053.155	45.312		
win=7, size=100					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-4.7630	1.2888	3662.9735	-3.696	<.001
diff_cosine	6.2542	2.5855	3561.5578	2.419	0.016
isi	-0.8312	1.2865	3732.1296	-0.646	0.518240
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	0	0		
target	(Intercept)	10.74	3.277		
Residual		2056.53	45.349		
win=7, size=150					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-4.0033	1.2851	3678.3956	-3.115	0.002
diff_cosine	8.9057	2.7455	3548.4574	3.244	0.001
isi	-0.7425	1.2866	3735.7114	-0.577	0.564
Random Structure	Name	Variance	Std.Dev.		
prime	(Intercept)	3.379e-06	0.001838		
target	(Intercept)	7.918e+00	2.813907		
Residual		2.060e+03	45.387624		

Lexical Decision data					
Default model (win=5, size=100)					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-18.272	1.688	4325.152	-10.826	<.001
diff_cosine	20.897	3.435	3075.541	6.084	<.001
isi1050	6.61	1.675	3072.477	3.946	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	32.73	5.721		
target	(Intercept)	65.08	8.067		
Residual		4213.42	64.911		
win=3, size=50					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-19.397	1.636	4424.068	-11.854	<.001
diff_cosine	16.371	3.071	3087.427	5.331	<.001
isi1050	6.383	1.679	3079.172	3.802	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	44.14	6.644		
target	(Intercept)	44.08	6.639		
Residual		4239.61	65.112		
win=3, size=100					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-18.313	1.665	4342.549	-10.998	<.001

diff_cosine	21.417	3.474	3069.667	6.165	<.001
isi1050	6.343	1.68	3065.82	3.776	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	37.97	6.162		
target	(Intercept)	48.97	6.998		
Residual		4224.78	64.998		
win=3, size=150					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-18.592	1.674	4326.78	-11.108	<.001
diff_cosine	21.902	3.725	3064.608	5.879	<.001
isi1050	6.496	1.68	3059.921	3.866	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	45.23	6.726		
target	(Intercept)	42.19	6.496		
Residual		4222.28	64.979		
win=5, size=50					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-18.767	1.657	4383.244	-11.326	<.001
diff_cosine	17.693	3.051	3083.764	5.799	<.001
isi1050	6.57	1.68	3075.758	3.912	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	32.31	5.684		
target	(Intercept)	47.77	6.912		
Residual		4240.87	65.122		
win=5, size=150					
Predictor	**Estimate**	**S.E.**	**df**	**t**	**p**

(Target/Control variable)					
(Intercept)	-18.398	1.697	4286.832	-10.844	<.001
diff_cosine	21.951	3.712	3059.534	5.913	<.001
isi1050	6.783	1.681	3056.978	4.035	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	42.81	6.543		
target	(Intercept)	49.89	7.063		
Residual		4219.23	64.956		
win=7, size=50					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-18.716	1.665	4407.067	-11.24	<.001
diff_cosine	17.338	3.043	3078.506	5.698	<.001
isi1050	6.161	1.68	3074.895	3.668	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	42.81	6.543		
target	(Intercept)	49.89	7.063		
Residual		4219.23	64.956		
win=7, size=100					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-18.143	1.695	4312.207	-10.705	<.001
diff_cosine	20.578	3.439	3071.519	5.984	<.001
isi1050	6.284	1.677	3072.699	3.748	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	30.61	5.533		
target	(Intercept)	58.79	7.667		

Residual		4221.21	64.971		
win=7, size=150					
Predictor (Target/Control variable)	**Estimate**	**S.E.**	**df**	**t**	**p**
(Intercept)	-18.11	1.698	4323.007	-10.665	<.001
diff_cosine	22.645	3.67	3075.249	6.17	<.001
isi1050	6.622	1.677	3074.563	3.949	<.001
Random Structure	**Name**	**Variance**	**Std.Dev.**		
prime	(Intercept)	33.63	5.799		
target	(Intercept)	60.36	7.769		
Residual		4221.74	64.975		