Modelling Reading Times in Bilingual Sentence Comprehension

Stefan L. Frank (s.frank@let.ru.nl)
Centre for Language Studies, Radboud University Nijmegen, The Netherlands

Abstract

Relatively little is know about the interaction between a bilingual's two languages beyond the word level. This paper investigates the issue by comparing word reading times (RTs) in both L1 and L2 to quantitative predictions by statistical language models. Recurrent neural networks are trained on either a Dutch corpus, an English corpus, or the two corpora combined (i.e., the bilingual network treats the two languages as one). Next, estimates of word surprisal by the three models are compared to RTs by native Dutch speakers on L1 Dutch and L2 English sentences. The monolingual Dutch model accounts for RTs on Dutch better than the bilingual model. In contrast, the bilingual model outperforms the monolingual English model on English RTs. These findings suggest that sentence comprehension in L1 is not much affected by L2 knowledge, whereas L2 reading does show interference from L1.

Keywords: Bilingualism; sentence comprehension; recurrent neural networks; word surprisal; word reading time

Introduction

Reading time (RT) effects on interlingual homographs and cognates have revealed that L1 knowledge affects L2 reading (Duyck, Van Assche, Drieghe, & Hartsuiker, 2007) and, vice versa, L2 knowledge affects L1 reading (Van Assche, Duyck, Hartsuiker, & Diependaele, 2009). However, whether these effects are modulated by sentence context (rather than being merely lexical phenomena) is still controversial (Libben & Titone, 2009; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011).

RT on a word depends, among other things, on the word's occurrence probability given the sentence so far. More precisely, a positive correlation has been found between RT and the negative logarithm of word probability, a value know as the word's *surprisal* (Fernandez Monsalve et al., 2012; Smith & Levy, 2013). Word surprisal can be estimated by statistical language models that are trained on large text corpora. So far, such work has only made use of models that process a single language (predominantly English) but if a bilingual's two languages influence each other during reading, bilingual (as opposed to monolingual) language models may provide a more accurate account of bilingual reading behaviour.

Modelling bilingual sentence processing

When recurrent neural networks (RNNs) are applied as statistical language models, their surprisal estimates regularly outperform those from other model types in predicting RTs (Frank & Bod, 2011; Frank & Thompson, 2012) as well as N400 size (Frank, Otten, Galli, & Vigliocco, 2013). Moreover, RNNs provide a straightforward account of how two languages may be combined into a single system, as the network's hidden layer can be activated by word input from either language without receiving any (explicit) information about language identity. French (1998) presents an early example of such a bilingual RNN, trained on two artificial miniature languages that were modelled on French and English. Since the current objective is to accurately estimate surprisal values for words from experimental stimuli or naturally occurring sentences, an RNN implementation is required that allows for training on large corpora of natural text. The highly efficient implementation by Mikolov, Deoras, Povey, Burget, and Černocký (2011) is well suited to this purpose.

Three RNNs were trained: one on a Dutch corpus, one on an English corpus, and one on the two corpora combined. Hence, there are two monolingual networks (henceforth, $\text{RNN}_{\text{Dutch}}$ and $\text{RNN}_{\text{English}}$) and one bilingual network (RNN_{bi}). Dutch training data came from a part of the Corpus of Web (Schäfer & Bildhauer, 2012; 5.8M sentences, 107M word tokens, 314K word types) and English data was taken from the British National Corpus (4.5M sentences, 87M word tokens, 182K word types). The three RNNs are architecturally identical, except for their number of input and output nodes which must match the number of word types in the training corpus. Hence, the only thing that makes a network Dutch, English, or bilingual is the language(s) it is trained on.

The RNNs embody two extreme views on bilingual processing: The monolingual models allow no effect of the other language whatsoever, whereas the bilingual model treats the two languages as one. Most likely, bilingual sentence comprehension falls somewhere in between these two poles. Fitting surprisal to RT should reveal which of the two extreme positions is most like bilingual reading. To the extent that bilinguals are affected by the language not currently being used, surprisal estimates by RNN_{bi} should fit the RT data better than surprisals from a monolingual RNN.

Results and conclusion

Surprisal values were obtained on each word of the 56 filler (i.e., non-target) sentences from a study by Frank, Trompenaars, and Vasishth (2014), who collected self-paced RTs from 46 native Dutch speakers tested in either Dutch ($N = 24$) or English ($N = 22$). $\text{RNN}_{\text{Dutch}}$ processed Dutch sentences, $\text{RNN}_{\text{English}}$ processed English, and RNN_{bi} processed both, yielding four sets of surprisal estimates. A significant amount of variance in RTs was accounted for by each set of surprisals (all $p < .0001$ in a linear mixed-effects regression analysis) over and above word length and word log-frequency.

The main question of interest is whether the monolingual RNNs' surprisal fits the data better or worse than surprisal from RNN_{bi}. Hence, we compare the fit to RTs of two regression models that differ only in the source of their surprisal values: one includes surprisal estimates by a monolingual RNN (i.e., $\text{RNN}_{\text{Dutch}}$ for Dutch; $\text{RNN}_{\text{English}}$ for English) and the other takes RNN_{bi}'s surprisals (on either Dutch or

Table 1: Model comparison results for monolingual versus bilingual RNN. Du. = Dutch; Eng. = English.

Lang.	Participants N	L1	L2	BF	$P(H_{\text{mono}})$
Du.	24	Du.	Eng.	23.2	.96
Eng.	22	Du.	Eng.	2.7×10^{-3}	.00
Eng.	20	Eng.	none	6.9×10^{4}	1.00
Eng.	20	Eng.	mixed	27.1	.96

English sentences). This is a comparison between two non-nested regression models, for which we take the approach advocated by Wagenmakers (2007): The difference between the two models' Bayesian Information Criterion gives rise to an estimate of Bayes Factor (*BF*) for the comparison between the two hypotheses (i.e., H_{mono} versus H_{bi}: the monolingual/bilingual RNN fits the data best). Assuming equal prior probabilities of H_{mono} and H_{bi}, *BF* is then used to obtain the probability of H_{mono} ($P(H_{\text{mono}})$, or, equivalently, $1 - P(H_{\text{bi}})$).

The first two rows of Table 1 show the estimated *BF* as well as $P(H_{\text{mono}})$ for the comparison between RNN_{bi} and either $\text{RNN}_{\text{Dutch}}$ (for RTs on Dutch sentences) or $\text{RNN}_{\text{English}}$ (for English). The RTs on L1 Dutch are predicted more accurately by $\text{RNN}_{\text{Dutch}}$ than by RNN_{bi}, whereas data on L2 English are predicted more accurately by RNN_{bi} than by $\text{RNN}_{\text{English}}$. This suggests that the Dutch native participants are not much affected by their L2 English when reading in Dutch, whereas their L1 does affect their reading process in English.

If the results for English really do reflect successful RNN modelling of L1 intrusion in L2 reading, then $\text{RNN}_{\text{English}}$ should outperform RNN_{bi} when fitting RTs from participants who do not speak Dutch. The UCL corpus (Frank, Monsalve, Thompson, & Vigliocco, 2013) provides eye-tracking data of native English speakers (with no knowledge of Dutch) reading 205 sentences that were randomly selected from unpublished novels. The bottom two rows of Table 1 show the results for monolingual and bilingual participants separately. As expected, both groups' RTs were predicted more accurately by $\text{RNN}_{\text{English}}$ than by RNN_{bi}. Hence, the Dutch/English bilingual RNN's success appears to depend on the data coming from Dutch/English bilingual readers.

This paper presented the first statistical language model that can process natural sentences from two different languages. A comparison between its word surprisal estimates and human RT measures provides evidence that L1 knowledge affects L2 comprehension beyond the word level, whereas the reverse may not be the case.

References

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective access. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *33*, 663-679.

Fernandez Monsalve, I., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 398–408). Avignon, France: ACL.

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, *22*, 829–834.

Frank, S. L., Monsalve, I. F., Thompson, R. L., & Vigliocco, G. (2013). Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, *45*, 1182–1190.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). Word surprisal predicts N400 amplitude during reading. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 878–883). Sofia, Bulgaria: ACL.

Frank, S. L., & Thompson, R. L. (2012). Early effects of word surprisal on pupil size during reading. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1554–1559). Austin, TX: Cognitive Science Society.

Frank, S. L., Trompenaars, T., & Vasishth, S. (2014). Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Manuscript submitted for publication*.

French, R. M. (1998). A simple recurrent network model of bilingual memory. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 368–373). Mahwah, NJ: Erlbaum.

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 381–390.

Mikolov, T., Deoras, A., Povey, D., Burget, L., & Černocký, J. (2011). Strategies for training large scale neural network language models. In *IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 196–201).

Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 486–493). Istanbul: ELRA.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302–319.

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*, 88–107.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*, *20*, 923–927.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779–804.