The processing of collocations by native and non-native speakers of English: Evidence from ERP studies

Jennifer Hughes (Lancaster University, UK)

In this presentation, I present the results of two experiments which combine methods from corpus linguistics and psycholinguistics in order to find out how the brain responds to encountering adjective-noun bigrams which form strong collocations (e.g. *clinical trials*) compared to adjective-noun bigrams which are semantically plausible but do not form collocations (e.g. *clinical devices*). In Experiment 1, the stimuli were presented to 16 native speakers of English; in Experiment 2, the same stimuli were presented to 16 non-native speakers of English (all native speakers of Mandarin Chinese).

The 15 collocational bigrams used in these experiments were selected because of their high forward transition probability. This is calculated by dividing the number of times the bigram X-then-Y occurs in the written BNC by the number of times X occurs in the written BNC altogether (McEnery and Hardie 2012:195). For each collocational bigram, I manufactured a non-collocational bigram that was absent from the BNC. The nouns in both conditions were matched for frequency and length, and all bigrams were embedded into corpus-derived sentences. These sentences were edited to ensure that they were semantically coherent as standalone units (Spöttl and McCarthy 2004:197) and that the sentence fragment preceding the experimental bigram was identical in both conditions. I also ensured that the preceding contexts created an equally "low contextual constraint" for the bigrams in each condition (Millar 2010:108).

The sentences were presented to participants at a rate of 500 ms per word (including a 200 ms interstimulus interval). Each participant wore a headcap containing 64 scalp electrodes, and these electrodes recorded some of the electrical activity of the brain while the participant was reading the stimuli. This method of "measuring electrical potentials in the brain by placing electrodes across the scalp" (Harley 2008) is known as EEG, or *electroencephalography*. One particular type of EEG study involves the measurement of ERPs, or *event-related potentials*, which are defined as "the momentary changes in electrical activity of the brain when a particular stimulus is presented to a person" (Ashcraft and Radvansky 2010:61). The ERP technique was used in the experiments discussed in this presentation, where the brain response was time-locked to the second word of each experimental bigram. These were then averaged across trials to isolate the experimental effect.

Previous studies have used the ERP technique to investigate semantic and syntactic processing. For instance, Kutas and Hillyard (1980) discovered that reading a semantic error elicits an N400, i.e. a negative voltage deflection occurring 400 ms after the onset of the error. Moreover, Osterhout and Holcomb (1992) found that reading a syntactic error elicits a P600, i.e. a positive voltage deflection that occurs 600 ms after the onset of the error. However, very few studies have used the ERP technique to investigate the processing of collocational errors, and those that *have* used ERPs for this purpose have tended to focus on idioms or other multi-word expressions. For instance, Molinaro and Cerrairas (2010:179-180) explicitly state that they focus on "idioms or clichés", and Molinaro et al. (2013:124) state that their stimuli is comprised of "multi-word expressions". The experiments discussed in this presentation therefore provide a unique contribution to the study of collocation by focusing on collocations that do not have clear beginning and end points and are not necessarily stored holistically.

The results of these experiments show that, for both native speakers and non-native speakers of English, reading the second word of a non-collocational bigram elicits an N400. It could be argued that this N400 is functionally distinct from the N400 discovered by Kutas and Hillyard (1980). While the N400 elicited in response to reading semantic errors typically has a

central-parietal scalp distribution, the results of my experiments show that the N400 elicited in response to reading collocational errors has an anterior-central scalp distribution. This replicates the results of my pilot study which was conducted on a different group of native English speakers.

From these experiments I can conclude that, for both native and non-native speakers of English, reading non-collocational bigrams elicits a quantitively distinct brain response compared to reading collocational bigrams. Interestingly, my results show that the N400 response is even larger for the non-native speakers. This finding is discussed in relation to the levels of prior exposure that both participant groups might have had to the bigrams outside the context of the experiments. Moreover, I discuss how the findings from these ERP studies provide evidence in support of a network model of language processing, whereby collocations are represented in the brain as transitions across a network.

References

Ashcraft, M. H. and Radvansky, G. A. (2010). *Cognition* (5th edn.). London: Pearson.

Harley, T. A. (2008). *The psychology of language: From data to theory* (3rd edn.). New York: Psychology Press.

Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427):203-205.

McEnery, T. and Hardie, A. (2012). *Corpus linguistics: Methods, theory and practice*. Cambridge: Cambridge University Press.

Millar, N. (2010). The processing of learner collocations (Unpublished doctoral thesis). Lancaster University, Lancaster.

Molinaro, N. and Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*(3):176-190.

Molinaro, N., Barraza, P. and Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, *72*:120-132.

Osterhout, L. and Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Language and Cognitive Processes*, *8*(4):785-806.

Spöttl, C. and McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (Ed.). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins, pp. 191-226.