What can MINERVA2 tell us about *killing hope*? Investigating L2 Collocational Processing as Memory Retrieval

Collocations are semi-productive word combinations wherein one word is used literally and one other figuratively. They are characterized by an arbitrary restriction on substitution (e.g., *kill hope*, *#murder hope*) and are notoriously difficult for L2 speakers to acquire and use. This difficulty is clearly reflected in processing and can be traced to complex interactions between *frequency* (exposure to L2 input), *congruency* (availability in the L1), and *semantic transparency* (degree to which meaning can be inferred from the constituent words). Research shows that collocations incur a processing cost over productive combinations (e.g., *eat cake*). To put this into perspective, fully opaque and non-productive idioms (e.g., *break the ice*) enjoy a processing advantage over productive combinations. Evidently, collocations not only possess idiosyncratic meanings, but also exhibit idiosyncratic processing trends, presenting a curious case for psycholinguistics, and language acquisition at large. However, despite increasing interest among researchers in Applied Psycholinguistics, little is known about the underlying mechanisms of collocational processing that can account for this idiosyncrasy.

The overarching aim of my thesis is to find a suitable explanatory model for L2 collocational processing. Based on evidence for analogical processes in L2 idiom and metaphor comprehension, I explore analogy as a plausible mechanism for collocations. Therefore, as an initial foray, this study attempts to explain trends in L2 collocational processing from the standpoint of memory retrieval—the first step in analogical processing. We parametrically explore the assumptions under which retrieval would be sufficient to explain the observed patterns of L2 learners' collocational processing. We use MINERVA2, a frequency-based, global-matching memory model, to simulate reaction times and compare them to data from 99 L1 and 230 L2 (L1 Portuguese) English speakers involving free combinations (e.g., eat cake ↔ 'comer bolos'), congruent (e.g., read minds ↔ 'ler mentes'), incongruent collocations (e.g., kick habits, no equivalent translation in Portuguese), and nonsense baselines (e.g., *read cakes*).

We make assumptions about: (i) how similar items are encoded in memory, and (ii) their sensitivity to frequency. Using DistilBERT models trained on English and Portuguese corpora, we extract two vector representations for each item capturing the semantics of the word combination in English, and its literal Portuguese translation. This allows us to model congruency and semantic transparency. We populate the MINERVA2 memory based on phrasal frequency in English, Portuguese, and a weighted (60-40) mix of Portuguese and English frequencies to model an intermediate-advanced L2 English learner. We simulate memory retrieval patterns from both languages under the various frequency conditions. Under the assumptions that the L2 lexicon develops with respect to the L1 and the L2 lexicon is sensitive to L1 frequencies, we find that MINERVA2 can neatly account for processing trends in both L1 and L2 collocational processing (see figures 1 and 2). Based on this, I would like to brainstorm with the audience about manipulating frequency conditions to simulate levels of L2 proficiency and model its effects on collocational processing.

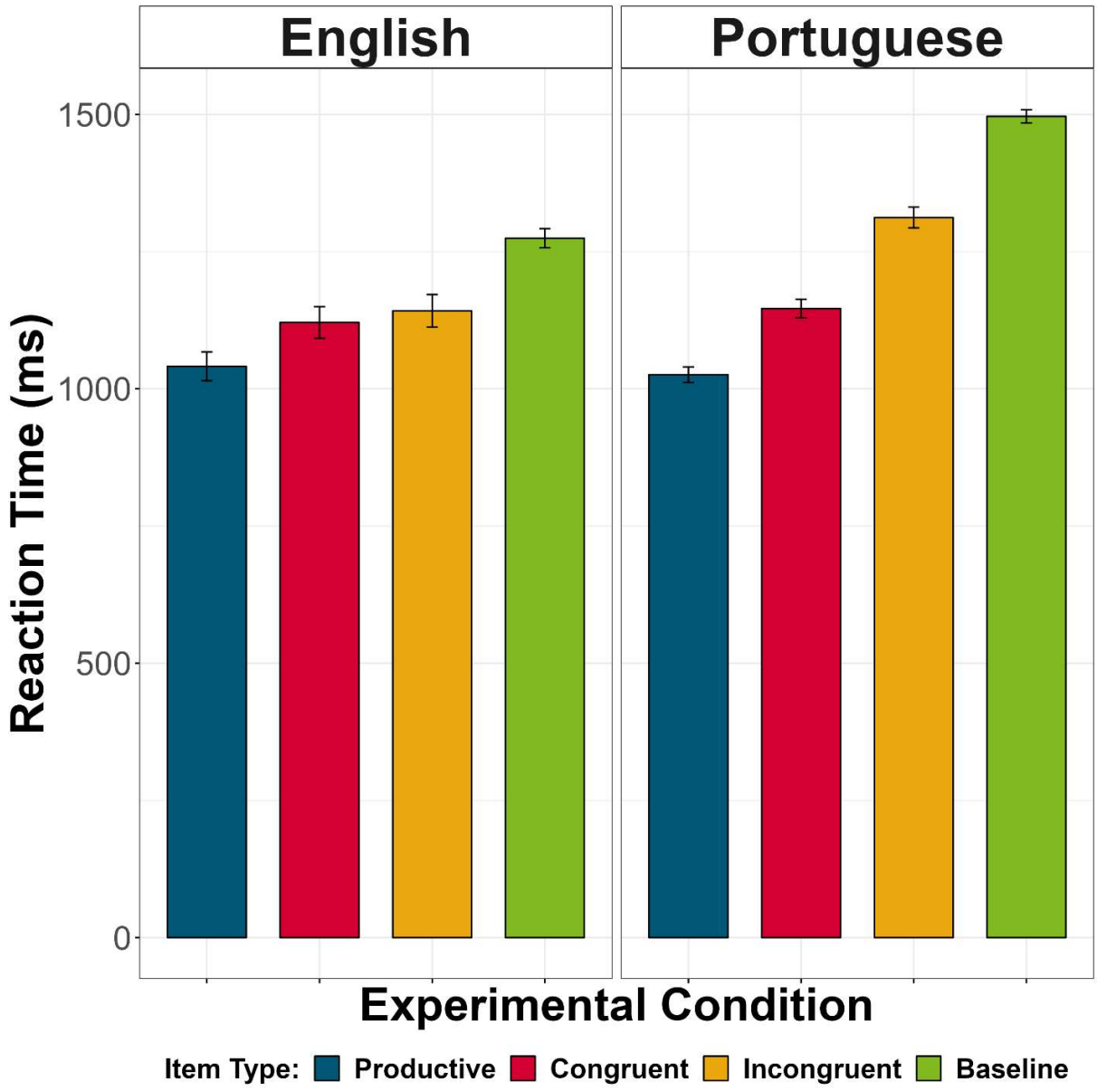

Figure 1: The mean reaction times (RTs) for L1 and L2 judgements. Error bars indicate bootstrapped confidence intervals.

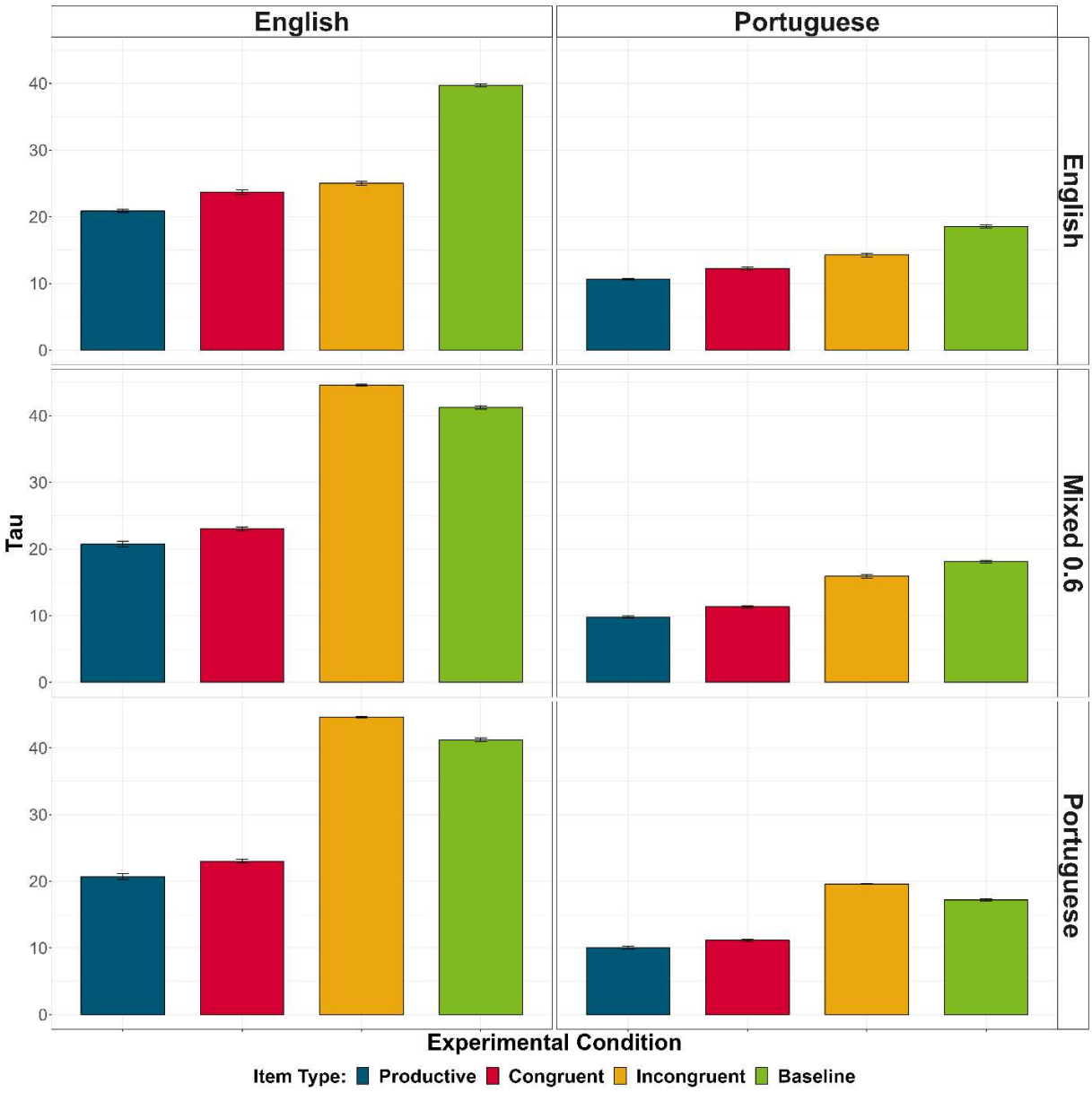

Figure 2: Simulated reaction times (tau) for all experimental conditions. Mixed 0.6 indicates the 60-40 frequency condition.