

Familiar statements are easier to interpret than unfamiliar statements in general, and these differences could influence the baseline correction region used for the analysis of the potentials elicited to the markers (see Urbach & Kutas, 2006 for detailed discussion of issues with baseline correction in ERP research). Baseline correction involves subtracting the mean amplitude in the baseline region from the post stimulus amplitudes. This subtraction shifts the entire waveform down or up by some constant. Thus, if there is a negative effect immediately preceding a critical word, centering the waveforms in the interval (i.e., baseline interval) must shift the more negative condition more positive everywhere following the interval (i.e., baseline). Likewise centering a condition that was more positive in this interval shifts this condition more negatively everywhere following the interval.

We examined possible differences in the amplitudes of familiar and unfamiliar proverbs that existed before the markers by creating epochs that extended from 200 ms before the beginning of each proverb to 500 ms after the onset of the markers (-200 to 4000 ms). We then conducted the same ANOVA reported in the manuscript (i.e., for N400/LP regions) on mean amplitudes in the 100 ms prior to the onset of the markers. As expected, there was only a significant effect of familiarity, $F(1,29) = 7.35$, $p < .01$, $\eta^2_p = .19$. Amplitudes were more negative for unfamiliar than familiar proverbs. The Figure below shows the familiarity effect at a central electrode located on the midline (CZ). This figure illustrates there is a larger N400 produced to the last word of unfamiliar than familiar proverbs. Critically, this N400 region includes the baseline interval used for the single word averages to the markers.

As a consequence of these pre-stimulus baseline differences, we examined the differences *between* the two markers when they followed each type of proverb as opposed to contrasting when the same marker was preceded by the two types of proverbs. Of course, this

means we chose to contrast brain potentials to different words. Note, however, that Ferretti et al. (2020) demonstrated that participants did not show differences in interpretation difficulty for *literally* and *figuratively* when they appeared at the beginning of sentences, and although *literally* has a higher lexical frequency than *figuratively* (14.49 vs 0.76 per million words; Brysbaert & New, 2009), *literally* produced the largest N400s in the present study.

Another consequence of the baseline issue is the interpretation of the direction of the main effects of familiarity in our analysis for the markers. Although the magnitude of the differences in familiarity effects reflect differences between the markers, the direction of the effect has been shifted in the opposite direction as a result of the amplitude differences in the baseline interval. For example, although our analysis shows that familiar proverbs had more negative amplitudes than unfamiliar proverbs in the N400 region, the actual effect is in the opposite direction.

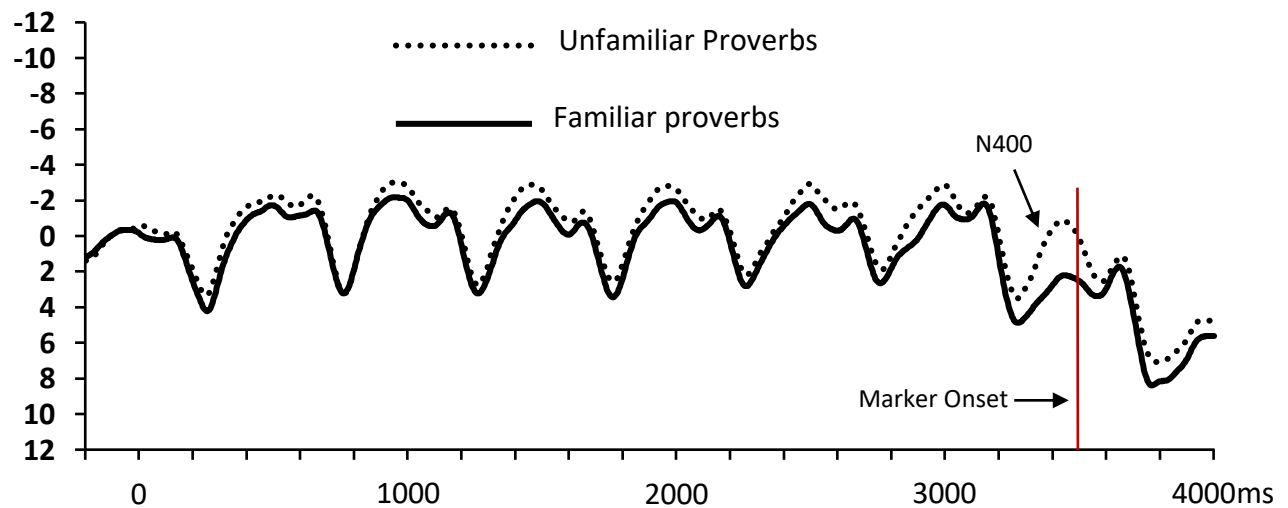


Figure 1: Mean amplitudes for familiar and unfamiliar proverbs that are timelocked to the first word of the proverbs and extending through to 500 ms after the onset of the marker.