The Effects of Word Frequency and Contextual Diversity on Word Processing during Reading across ERP and Eye-tracking Measures

Patrick Plummer (Howard University) & Charles Perfetti (University of Pittsburgh)
patrick.plummer@howard.edu

Recent psycholinguistic research has suggested that the distributional properties of words across distinct contexts exert an influence on word processing during language comprehension. The current study examined the effects of word frequency (how often a given word appears) and contextual diversity (the number of contexts in which a given word appears) using eye tracking and electrophysiological measures. Using measures obtained from the SUBTLEX corpus, target word pairs were constructed to contrast either contextual diversity or word frequency, while matching other important lexical variables (e.g. word length, concreteness, and number of senses). Half of the pairs varied in word frequency while matching in contextual diversity and half varied in contextual diversity while matching in word frequency. In both experiments, participants read sentences where one member of the target word pair was embedded. Each sentence frame was designed to be semantically neutral prior to the target word position. In Experiment 1, participants read the sentences for comprehension while their eye movements were being monitored; eye movement measures associated with word recognition and integration were analyzed. In Experiment 2, participants read sentences for comprehension while electrophysiological activity was monitored from their scalp via EEG recording; ERP components associated with relatively early word processing (i.e. N170 and P2) as well as lexical-semantic processing and integration (i.e. N400) were analyzed.

In Experiment 1, the effect of contextual diversity superseded the standard word frequency effect. First-pass and later reading times for high frequency words did not differ from low frequency words when controlling for contextual diversity. On the other hand, reading times for high contextual diversity words were significantly shorter than reading times for low contextual diversity words when controlling for word frequency. In Experiment 2, when controlling for contextual diversity, no significant effects of word frequency were observed for any ERP components of interest. In contrast, across midline electrode clusters (Fz, Cz, Pz), smaller P2 components and larger N400 components were observed for high contextual diversity words when compared to low contextual diversity words. This pattern of results conflicts with the typical word frequency effect on ERP components wherein high frequency words elicit smaller N400 components than low frequency words. Thus, the results revealed that the effects of contextual diversity on eye movement control during reading (i.e. analogous to typical word frequency effects) were somewhat at odds with the effects of contextual diversity on electrophysiological responses during reading (i.e. opposite of typical word frequency effects).

Overall, the results showed that the typical word frequency effect in ERP and eye-tracking measures disappeared when controlling for contextual diversity; whereas, contextual diversity effects persisted when accounting for word frequency. Interestingly, the larger N400 components for high contextual diversity words suggests that greater integration difficulty or broader semantic activation was associated with processing words that appear across a many distinct linguistic contexts. The pattern of results in Experiment 2 suggests that contextual diversity has an influence on word processing regardless of word frequency. However, the fact that the ERP effects of CD differ from standard ERP effects of word frequency when the two variables are controlled, suggests that contextual diversity is not simply a “better measure of frequency.” In general, the findings provide further evidence that language users’ experience with the contexts in which a word typically appears is an important aspect of lexical representation and seems to hold more psychological relevance than mere frequency of exposure to specific lexical items.

Experiment 1 Stimuli

	Word Frequency Contrast Group	Contextual Diversity Contrast Group
High	WF = 3.46 (0.7) CD = 2.92 (0.5)	WF = 3.25 (0.5) CD = 3.01 (0.5)
Low	WF = 3.17 (0.6) CD = 2.91 (0.5)	WF = 3.24 (0.5) CD = 2.75 (0.4)

Experiment 1
First-pass reading times

	WF Contrast	CD Contrast
High	253 (49)	260 (52)
Low	252 (55)	276 (55)

Group/Condition Interaction -
$F(1,53) = 5.39$, $p < .05$, $\eta_p^2 = 0.092$

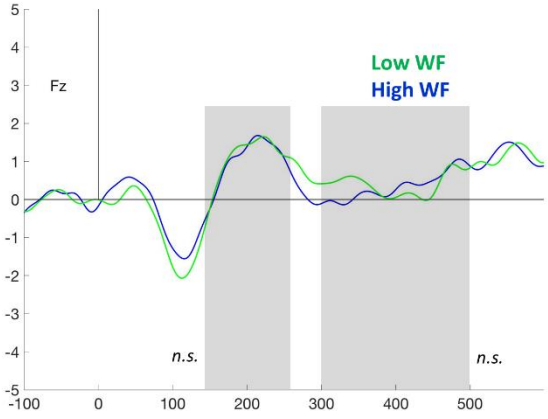
Experiment 2 Stimuli

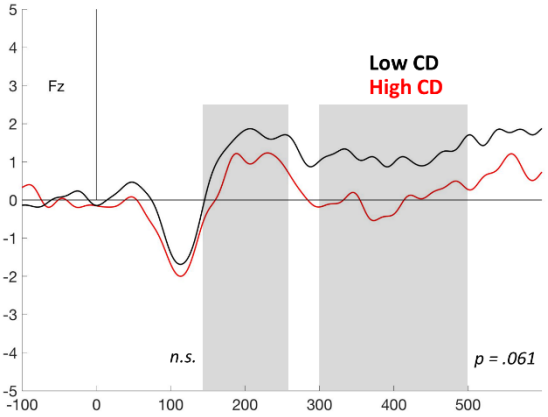
	Word Frequency Contrast Group	Contextual Diversity Contrast Group
High	WF = 3.24 (0.50) CD = 2.78 (0.5)	WF = 2.86 (0.6) CD = 2.68 (0.6)
Low	WF = 2.99 (0.4) CD = 2.78 (0.5)	WF = 2.84 (0.6) CD = 2.43 (0.6)

Stimuli tables show mean (log-transformed) CD and WF for high and low items from both contrast groups. Standard deviations in parenthesis. Each group contained 20 target word pairs for experiment 1 and 35 word pairs for experiment 2.

Experiment 2 ERP Results – Combined midline cluster ANOVA (Fz, Cz, Pz) showed high CD items led to smaller P2 and larger N400 components than Low CD items (*p*'*s* < .01). No differences for WF contrast items on any standard ERP components.

Midline Frontal





Midline Central