



Semantic diversity, frequency and the development of lexical quality in children's word reading

Yaling Hsiao, Kate Nation

Department of Experimental Psychology, University of Oxford, Oxford, UK



ARTICLE INFO

Keywords:

Semantic diversity
Contextual diversity
Frequency effect
Lexical quality
Reading development

ABSTRACT

Frequency exerts a powerful influence on lexical processing but it is possible that at least part of its effect is caused by high frequency words being experienced in more diverse contexts over an individual's language experience. To capture this variability, we applied Latent Semantic Analysis on a 35-million-word corpus of texts written for children, deriving a measure of semantic diversity that quantifies the similarity of all the contexts a word appears in. Across three experiments with 6–13-year-old children involving reading aloud and lexical decision, we found a main effect of semantic diversity: high diversity words were responded to faster and read more accurately than low diversity words. Frequency, document count and age of acquisition were also significant predictors of reading behaviour. These findings demonstrate that contextual variability contributes to word learning and the development of lexical quality, beyond the effect of frequency.

Introduction

Children who read more are better at reading words than children who read less. Estimates of print exposure – a proxy for how much an individual has read – account for unique variance in reading development and are associated with individual differences in orthographic and phonological processing (e.g. Cunningham & Stanovich, 1990; Mol & Bus, 2011). Why might this be? Reading is a skill and like all skills, practice is critical to becoming expert and for word-level reading, practice may be important in at least two distinct ways. First, reading practice allows basic skills to be honed and fine-tuned, promoting the development of reading fluency. In addition, however, an avid reader is likely to experience more words and a larger range of different language contexts than a less prolific reader. In this paper, we investigate whether this broad lexical experience influences the ease with which children read words.

Reading experience provides exposure to individual words, cumulatively adding to both type and token frequency. Word frequency is an item level variable that typically reflects the number of times a word appears in a corpus. In adults, the frequency effect is robust across a range of lexical tasks involving written words, with high frequency words enjoying a processing advantage (Balota & Chumbley, 1984; Brysbaert, Warriner, & Kuperman, 2014; Monsell, 1991; Rayner & Duffy, 1986) (for review, see Brysbaert, Mander, & Keuleers, 2018). In children too, estimates of word frequency influence how long it takes to read a word, or make a lexical decision to it (Joseph, Nation, &

Liversedge, 2013; Schmalz, Marinus, & Castles, 2013). Models of skilled word recognition vary in how they handle the frequency effect (e.g. Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Murray & Forster, 2004; Norris, 2006; Plaut, McClelland, Seidenberg, & Patterson, 1996) but centre on the idea that variations in frequency reflect differences in experience, with repeated exposure to a word influencing its accessibility, either by changing the recognition threshold (e.g. McClelland & Rumelhart, 1981), or the weights between nodes in distributed models (e.g. Seidenberg & McClelland, 1989). Surprisingly, given the ubiquity of the frequency effect, there has been little discussion of how reading experience shapes lexical representations through development so as to influence reading behaviour.

In line with these theoretical accounts, frequency might influence the development of lexical quality via the principle of repetition. On this view, words higher in frequency (and therefore experienced more often) become more strongly represented in memory over time and leading them to be processed more efficiently than words lower in frequency (and therefore experienced less often). A quite different theoretical account is that frequency influences lexical processing via the quality or contextual nature of encounters with each word, not just the number of encounters. In natural reading, words are rarely encountered in isolation: they occur in sentences, paragraphs and texts. Does the linguistic context in which a word has occurred across previous encounters matter for how that word is subsequently processed? The lexical legacy hypothesis (Nation, 2017) suggests it might, as it is the substrate from which knowledge about a word builds. On this view,

E-mail addresses: yaling.hsiao@psy.ox.ac.uk (Y. Hsiao), kate.nation@psy.ox.ac.uk (K. Nation).

<https://doi.org/10.1016/j.jml.2018.08.005>

Received 22 February 2018; Received in revised form 14 August 2018

0749-596X/© 2018 Elsevier Inc. All rights reserved.

reading experience provides many different contexts and episodes which over time result in a complex database about a word, its connections to other words and its lexical history within an individual's experience. In turn, these rich and diverse encounters bring about local variation in lexical quality (Perfetti, 2007) at the word level: a legacy that is measurable during word reading behaviour, even as skilled readers process words in a laboratory task such as lexical decision.

Several lines of evidence suggest that variation in contextual experience with words influences how easily they are processed in tasks such as lexical decision. McDonald and Shillcock (2001) devised a new variable, *contextual distinctiveness*, which captured the local lexical environment in which words co-occur across a corpus. They defined contextual distinctiveness as the relative entropy between the posterior distribution (the distribution of words occurring in a ten-word window around a target word) and the prior distribution (the expected distribution of those words when the target word is not taken into account). Contextual distinctiveness was a better predictor of lexical decision latencies than word frequency, supporting the view that distributional statistics that reflect the lexical environment a word has been experienced in have a role to play in theoretical accounts of visual word recognition (see Baayen, 2010 for extended discussion).

A number of more recent studies (for review, see Jones, Dye, & Johns, 2017) offer support to the general approach introduced by McDonald and Shillcock (2001). Adelman, Brown and Quesada (2006) found that contextual diversity (indexed in their study as the number of unique documents a word appears in across a corpus) not only predicted lexical decision and naming latency, it eliminated any effect of word frequency. Variations in document count are also associated with word reading in sentence processing, as revealed by analysis of eye movements (Plummer, Perea, & Rayner, 2014). One issue with these findings is that contextual diversity as indexed by document count is highly correlated with word frequency and it might be that it is simply a better measure of frequency than frequency itself (e.g. Brysbaert & New, 2009). It is also important to note that document count does not take the content of the contexts into account. This is an important point to consider, if we are to understand the theoretical reasons for why document count might influence lexical processing, beyond frequency.

Semantic diversity (or its reciprocal, *semantic distinctiveness*) is a variable that explicitly captures the similarity in content of different contextual experiences of a word (e.g. Hoffman, Lambon Ralph, & Rogers, 2013; Johns, Gruenfelder, Pisoni, & Jones, 2012; Jones et al., 2017; Jones, Johns, & Recchia, 2012). It is similar in spirit to McDonald and Shillcock's (2001) contextual distinctiveness variable, but rather than calculating contextual similarity via local lexical co-occurrence, similarity is quantified using Latent Semantic Analysis techniques. This produces a metric that captures context-dependent variation in similarity in a graded fashion. Semantic diversity predicts lexical decision and naming latency in adults, outperforming both word frequency and document count (Johns et al., 2012; Johns, Dye, & Jones, 2014, 2016), with words that appear in more semantically diverse contexts being processed more easily than less semantically diverse words. In this way, semantic diversity behaves similarly to polysemy, and indeed, the processing advantage for polysemous words in lexical decision might be related to the fact that polysemous words tend to be more semantically diverse (Azuma & Van Orden, 1997; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Rodd, Gaskell, & Marslen-Wilson, 2002).

Clearly, semantic diversity is associated with how easily skilled adult readers make a lexical decision response. Its basis, however, must stem from reading and language experience. Retuning to the lexical legacy hypothesis (Nation, 2017), a word's semantic diversity at any point in time can be thought of as the product of an individual's contextual experiences with that word and the opportunities for learning that are afforded by those experiences, culminating in variations in lexical quality that in turn govern item-level variation in lexical processing. This description chimes with Adelman et al.'s conclusion that "learning based models of reading cannot accommodate [Adelman

et al.'s] results unless they are modified so that learning mechanisms are sensitive to context, not frequency" (2006, p. 822). While there have been computational implementations of word learning from contextual experience (Hoffman & Woollams, 2015; Johns et al., 2014), it is striking that there is little relevant evidence from studies of children's reading. Given that semantic diversity is a variable that has its roots in learning and experience, developmental data are important.

There is, however, evidence to show document count influences on children's word reading. Perea, Soares and Comesaña (2013) calculated a contextual diversity variable from a children's reading corpus, using document count as their metric. They chose 60 words that varied orthogonally in contextual diversity and frequency and asked 4th Grade Portuguese children to make a lexical decision to them. Words high in contextual diversity were processed more quickly than frequency-matched low diversity words. There was, however, no effect of frequency: latencies to high vs. low frequency words did not differ when the two sets of words were matched for contextual diversity. These findings replicated in a second experiment that adopted a regression design, using a different sample of children and a different set of items.

By 4th grade then, it seems that children are sensitive to contextual factors afforded by reading experience. One limitation to these findings is that diversity was instantiated using document count, rather than semantic diversity. As noted above, document count is highly correlated with word frequency and does not capture the similarity of content between contexts. Sample size was small, with fewer than 30 children in each experiment, and Perea et al. only sampled 60 items. The children's corpus was also small in terms of the number and range of documents, comprising 3.2 million words taken from 171 elementary textbooks.

In our paper, we aimed to build on these findings. Most importantly, we devised a measure of semantic diversity to capture the semantic similarity in content across contexts using the *Oxford Children's Corpus* – a developmental corpus of children's written language. Based on evidence from adults, we predicted that this would be less tightly bound with word frequency than document count, and associated with variation in children's lexical decision and naming. We also asked whether frequency influences children's lexical processing. In adults, evidence suggests that word frequency matters less, once semantic diversity is taken into account. In children, however, frequency might be more important as repetition might be critical in a developing system. With this in mind, we also measured the children's reading level and extended the age range, allowing us to investigate whether semantic diversity and frequency have different effects at different levels of proficiency. Finally, we also considered age of acquisition, defined as the approximate age at which a word is learned. Like frequency, age of acquisition influences children's lexical processing (Coltheart, Laxon, & Keating, 1988). Its association with semantic diversity has not been investigated in relation to word reading, although there is evidence that in infancy, words that are acquired early are experienced in more diverse contexts and as a result become more associated with other known words (Hills, Maouene, Riordan, & Smith, 2010).

We begin by introducing the *Oxford Children's Corpus* and describing how semantic diversity was calculated before examining how variations in semantic diversity are associated with children's lexical processing in three different datasets.

Calculating semantic diversity

The Oxford Children's Corpus (OCC) is a dynamic and growing corpus, initiated in 2006 by Oxford University Press to guide the preparation of dictionaries for children. The version used in this paper contained over 35 million words and 12,000 documents, targeted at children aged 5–16 years old. Unlike some other children's corpora, the OCC is not restricted to curriculum materials and structured reading schemes. It also contains classic and modern children's fiction, non-fiction, textbooks, websites and magazines. Thus, it broadly samples a

Download English Version:

<https://daneshyari.com/en/article/9953006>

Download Persian Version:

<https://daneshyari.com/article/9953006>

[Daneshyari.com](https://daneshyari.com)