Journal of Memory and Language 87 (2016) 38–58

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Part and whole linguistic experience affect recognition memory for multiword sequences

Cassandra L. Jacobs [a,*], Gary S. Dell [a], Aaron S. Benjamin [a], Colin Bannard [b]

[a] University of Illinois at Urbana-Champaign, United States
[b] University of Liverpool, United Kingdom

ARTICLE INFO

Article history:
Received 17 January 2015
revision received 27 October 2015

Keywords:
Phrase frequency
Compositionality
Word frequency
Recognition memory

ABSTRACT

Low frequency words (like *wizard*) are better remembered in recognition memory than high frequency words like *tree*. Previously studied low frequency words are endorsed more often than high-frequency words, and unstudied low frequency lures attract fewer false alarms than high frequency lures. In order to evaluate whether repeated experience of phrases has the same effect as that of words, we tested whether infrequent combinations of words (like *psychic nephew*) are better recognized than frequent word combinations (like *alcoholic beverages*). In contrast to single words, people were more biased to endorse high-frequency phrases, but phrase frequency did not affect discrimination between studied and unstudied phrases. When high and low frequency nouns were embedded in adjective-noun phrases of equal frequency (e.g. *handsome wizard* and *premature tree*), people were better able to recognize phrases containing low frequency than high frequency nouns. Taken together, the high frequency phrase bias and the low frequency embedded-noun advantage suggest that the recognition of word sequences calls on prior experience with both the specific phrase and its component words.

Introduction

Researchers have carried out thousands of experiments in which word frequency is manipulated with the goal of understanding how words are processed, produced, and remembered. For the most part, this research demonstrates that low frequency words are less easily acquired, comprehended, and produced than more common words (see Ellis (2002) for a complete review). More recently, the question of whether multi-word sequences (or phrases) might exhibit frequency effects has been assessed. As with common words, high-frequency phrases are associated with benefits in reading time (Bannard, 2006; Smith & Levy, 2013), phrase decision reaction time (Arnon & Snider, 2010), greater fluency and speed of production (Arnon & Priva, 2013; Bannard & Matthews, 2008; Janssen & Barber, 2012) and recall memory (Tremblay & Baayen, 2010).

Phrase frequency effects are of interest because they tell us about the cognitive mechanisms implicated in the production and comprehension of word sequences. The findings cited above indicate that the combination matters. A phrase is not just a list of words. More importantly, these results are analogous to the discovery in morphology that people are sensitive to the frequency of whole words, and the inference that word processing involves some knowledge of the whole as well as of the component morphemes (Bien, Levelt, & Baayen, 2005). However there remain many questions about the exact nature of the mechanisms involved. There are two main issues that arise: compositionality and abstraction. In this paper, we

* Corresponding author at: 603 E. Daniel St., Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL 61820, United States.

E-mail address: cljacob2@illinois.edu (C.L. Jacobs).

http://dx.doi.org/10.1016/j.jml.2015.11.001
0749-596X/

present five recognition memory experiments that address these issues.

The compositionality issue concerns the representation of a phrase, by which we loosely mean the mental/neural codes implicated in producing and understanding it, and whether these codes are a predictable superset of the representational spaces involved in the production and comprehension of its parts. So, a person's knowledge of the phrase *red house* may be compositional, derived solely from their knowledge of its component words, *red* and *house*. If the phrase is not compositional, but instead holistic, a language user's representation of it might be largely separate from their representation of the component words. Phrases vary in the extent to which their meaning is predictable from their parts, with the meaning of *red house* being much more predictable than the meaning of *red herring*. A phrase with an unpredictable meaning therefore may seem to require a largely disjoint representation. It is also plausible that such representations might also be employed for more predictable phrases as well. Indeed, the discovery of phrase-frequency effects has occasionally been taken to indicate that the representation of phrases is holistic. However, while such results indicate that speakers do encode knowledge of the sequences, they do not address the question of whether combination-specific knowledge is utilized instead of or in addition to word knowledge when processing phrases.

The issue of abstraction concerns how we encode multiple instances of the same phrase. A phrase could be represented either as a collection of episodic memories, each containing a token of that phrase, or as a single abstract encoding of the type with an associated strength. In the episodic approach, the particular episodes in which a phrase is experienced are kept distinct, and effects of the phrase frequency would be attributed to the number of such episodes. In particular, any processing benefits that accrue to common phrases would be attributed to the greater availability of relevant memories to guide the processing (e.g. Goldinger, 2004; Hintzman, 1988). Alternatively, in the abstractionist approach, each phrase type is a single representation such as a node in a lexical-semantic network (e.g. MacKay, 1982). If *red house* had been experienced a number of times, a node would represent the phrase type, with its strength (e.g. resting level of activation) proportional to its frequency. Of course, the abstractionist approach does not deny the existence of episodic knowledge about phrases. It simply assumes that the abstraction exists in addition to episodic memories, and it is this abstraction that plays the major role in how the phrase is processed, rather than the episodes.

Some accounts of word and phrase frequency effects are neither clearly episodic nor explicitly abstractionist in the sense that they have a single node for each word or phrase. Multi-level connectionist models (e.g. Seidenberg & McClelland, 1989) occupy an interesting middle ground in this respect. Each experience changes the weights in a network (as with an episode) and yet these alterations are not stored separately, but rather are superimposed. The resulting superposition is somewhat like an abstraction, but it is not easily recognized as such and is certainly not a single node. A related class of models, the naive discrimination learning models (e.g. Baayen, Hendrix, & Ramscar, 2013; Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011), also lacks discrete episodes and explicit representations of abstract items. For example, one such model by Baayen et al. (2011) consists of an input layer of letters and letter pairs and an output layer of semantic features. The model learns input–output mapping for words or phrases by applying the Rescorla–Wagner (1972) equations to probabilistic information obtained from corpora. Even though it lacks explicit words or phrases, its behavior (e.g. mapping accuracy) reflects both word and phrase frequency.

As we noted, benefits for high frequency phrases have been clearly demonstrated in comprehension and production tasks, and in memory recall. In our studies, we turn to a different memory task in order to address phrase frequency from a new perspective: the yes–no recognition task. Importantly, the high frequency advantage apparent in linguistic tasks and recall is not evident in recognition memory; in fact, low frequency words are recognized better. We more easily pick out *panther* when it was studied and reject it when it was not studied than a higher frequency word like *cat*. That is, low frequency words attract more hits and fewer false alarms than high frequency words. This pair of results is one manifestation of a broader category of what are called mirror effects, effects in which a particular class of items or condition of study for a set of items leads to them being more easily discriminated (Glanzer & Adams, 1985). The mirror effect allows us to derive predictions about frequency effects for phrases in recognition, and thus examine the cognitive mechanisms implicated in their processing.

In the next section we review studies of word frequency in language processing and acquisition. Next, we discuss the degree to which the high frequency word advantage is reflected in larger sequences of linguistic units, such as multi-word sequences. Finally, we review the mirror effect in yes–no recognition memory and consider its implications for multi-word sequences.

The high frequency word advantage

High frequency words are easier to process than low frequency words. The language processing system is adaptive and thus learns to process more probable events with greater facility (Dell & Jacobs, 2015; Forster & Chambers, 1973; Jusczyk, 1997; Lively, Pisoni, & Goldinger, 1994; Saffran, Newport, & Aslin, 1996). For example, identification of high frequency words is more robust under both noisy (Howes, 1952) and clear (e.g. Forster & Chambers, 1973) conditions.

When reading words in text, reading times scale inversely with the logarithmic frequency of the word that is being read, with the most common words in the language being barely read at all or even skipped entirely (e.g. Demberg & Keller, 2008; Howes & Solomon, 1951; Rayner, 1998; Smith & Levy, 2013). When a text contains low frequency words, comprehension suffers (Diana & Reder, 2006; Freebody & Anderson, 1983; Marks, Doctorow, & Wittrock, 1974). In production, uncommon words are retrieved more slowly during picture naming (e.g. Oldfield & Wingfield, 1965) and produced less