Cognition 157 (2016) 384–402

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Abstract knowledge versus direct experience in processing of binomial expressions

Emily Morgan [a,b,*], Roger Levy [a,c]

[a] Department of Linguistics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0108, United States
[b] Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, United States
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, United States

ARTICLE INFO

Article history:
Received 13 May 2014
Revised 17 August 2016
Accepted 15 September 2016
Available online 21 October 2016

Keywords:
Binomial expressions
Word order
Sentence processing
Comprehension
Frequency

ABSTRACT

We ask whether word order preferences for *binomial expressions* of the form *A and B* (e.g. *bread and butter*) are driven by abstract linguistic knowledge of ordering constraints referencing the semantic, phonological, and lexical properties of the constituent words, or by prior direct experience with the specific items in questions. Using forced-choice and self-paced reading tasks, we demonstrate that online processing of never-before-seen binomials is influenced by abstract knowledge of ordering constraints, which we estimate with a probabilistic model. In contrast, online processing of highly frequent binomials is primarily driven by direct experience, which we estimate from corpus frequency counts. We propose a trade-off wherein processing of novel expressions relies upon abstract knowledge, while reliance upon direct experience increases with increased exposure to an expression. Our findings support theories of language processing in which both compositional generation and direct, holistic reuse of multi-word expressions play crucial roles.

1. Introduction

When we encounter common expressions like *I don't know* or *bread and butter*, do we process them word-by-word or do we treat them as holistic chunks? Research on sentence processing has largely focused on how single words are combined into larger utterances, but intuitively it seems that high frequency multi-word expressions might be processed holistically, even if they could in principle be treated compositionally. Recent research has thus questioned what possible sizes of combinatory units should be considered as the building blocks of sentence processing: Must all multi-word expressions be generated compositionally each time they are used, or can the mental lexicon contain holistic representations of some multi-word units?

The primary diagnostic for this question is whether the frequency of occurrence of multi-word expressions is predictive of their behavior in language processing. Such frequency effects are well documented at the level of individual words: more frequent words are faster to read (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Rayner, Sereno, & Raney, 1996), more likely to be skipped in reading (Rayner et al., 1996; Rayner & Well, 1996), and more susceptible to phonetic reduction (Bybee, 1999; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999). But do comparable frequency effects exist for multi-word expressions, when the frequency of their component words is controlled for? If the frequency of a given expression is being mentally stored, this implies that there is a mental representation of the expression as a whole. In contrast, if there are no frequency effects at the level of multi-word expressions, this is evidence against them having holistic representations akin to those of individual words.

A traditional view of grammar does not include holistic representations of multi-word expressions. According to this view, there is a strict separation between the individual words of a language and the rules for combining them. Pinker (2000), for example, describes a "traditional words-and-rules theory" in which "there are two tricks, words and rules. They work by different principles, are learned and used in different ways, and may even reside in different parts of the brain." (See also Ullman, 2001; Ullman et al., 2005.) One tenet of this theory is that forms which can be generated compositionally are not stored: for instance, in the case of the English past tense, irregular forms are stored, while regular forms are generated anew using the *-ed* suffix each time they are used (Pinker, 1991). It remains possible within this theory that some regular forms—particular extremely high frequency ones—may be stored as well, but this is not the general method for dealing with such forms. As Pinker (2000) explains, one key motivation

* Corresponding author at: Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, United States.
E-mail address: emily.morgan@tufts.edu (E. Morgan).

http://dx.doi.org/10.1016/j.cognition.2016.09.011
0010-0277/

for this theory is memory constraints on the representation of language knowledge: it is more efficient to store a single, widely applicable rule than to store each regular form individually.

In a similar vein, this theory predicts that multi-word expressions should not be stored holistically because they can be generated compositionally, except in the case of non-compositional exceptions such as idioms (Swinney & Cutler, 1979). Again, as with regularly inflected wordforms, some exceptions may exist, but the exponentially larger number of multi-word expressions with which people have experience makes it even less likely that these expressions would be stored holistically, given the motivating concern with storage efficiency. The words-and-rules theory thus does not predict that the processing of a multi-word expression will be affected by the frequency of the expression as a whole, though it can be affected by the frequencies of the individual words making up the expression.[1]

In contrast, there exists a growing movement of grammatical theories that do not draw a sharp distinction between the lexicon and the combinatory rules (e.g. Baayen, Milin, Durdevic, Hendrix, & Marelli, 2011; Bybee, 2001, 2006; Gahl & Yu, 2006; Goldberg, 2003; Hay & Bresnan, 2006; Johnson, 1997, 2006; Langacker, 1987; Pierrehumbert, 2000; van den Bosch & Daelemans, 2013). Rather than conceiving of rules as static entities dissociated from the lexicon, these *usage-based* approaches instead conceive of rules as dynamically generated generalizations over one's linguistic experience. In particular, many of these approaches (notably Bybee, 2001; Hay & Bresnan, 2006, among others) claim that people mentally store *exemplars*, or tokens of linguistic experience, which can be larger than single words. Language users then form generalizations from exemplars at multiple levels of granularity (e.g., morpheme, word, or phrase) simultaneously, and the resulting network of generalizations constitutes our grammatical knowledge. Single words and multi-word expressions are thus on an equal footing: both are possible units that can be inferred from exemplars, and frequencies of multi-word expressions are predicted to be stored and tracked just as frequencies of single words are.

Similar claims are made by exemplar-based computational models, which, like the exemplar-based grammatical theories, can incorporate combinatorial units of varying sizes from morphemes to sentences (e.g. Bod, 1998, 2008; Bod et al., 2003; Johnson, Griffiths, & Goldwater, 2007; O'Donnell, Snedeker, Tenenbaum, & Goodman, 2011; Pierrehumbert, 2000; Post & Gildea, 2013). Within these models, the process of learning a grammar is explicitly one of deciding what sizes of units are most applicable or probable to explain the available language data. Under the learned grammars, many utterances can be parsed in multiple ways, either as combinations of individual words, or as holistic expressions, or various combinations thereof.

Evidence for these usage-based theories in the domain of multi-word expressions comes in large part from previous demonstrations of phrase-level frequency effects. Bybee (2006) reviews numerous corpus analyses demonstrating that the frequency of multi-word expressions is predictive of phonological reduction, grammaticalization, and other properties of usage, with a focus on highly frequent expressions such as *I don't know* or *going to*. Frequency effects for multi-word expressions have also been demonstrated in a controlled experimental setting: in a phrasal-decision task (analogous to a lexical decision task), Arnon and Snider (2010) found that more frequent phrases—e.g. *Don't have to worry*—were judged to be sensible phrases of English faster than less frequent phrases matched for word and substring frequencies—e.g. *Don't have to wait*. They further demonstrate that these effects exist across a wide range of frequencies, not just at the highest end of the frequency spectrum. (For a comparable finding using phonetic duration in corpus data, see Arnon & Cohen Priva, 2013. Similar frequency effects have also been found in child language acquisition; see Bannard & Matthews, 2008.)

The exemplar-based approach also accords with more recent work on idioms, which challenges the traditional notion of idioms as strictly non-compositional. Gibbs (1990) and Nunberg, Sag, and Wasow (1994) argue that many idioms can be seen as conventionalized metaphoric extensions of their literal meanings, and thus need not be treated as exceptions to the prevailing rules. (Similarly, see Holsinger, 2013.) On the whole, we thus see a broad shift towards recognizing that many expressions reside in a grey zone between entirely compositional and entirely non-compositional, and furthermore that an expression may be conventionalized while still being at least somewhat compositional.

But there remain open questions regarding these exemplar-based approaches and the interpretation of frequency effects for multi-word expressions. One limitation in the work to date is that it is difficult to differentiate the effects of language experience per se from the effects of real-world knowledge. Bybee (2006), for example, stresses the importance of language experience:

> As is shown here, certain facets of linguistic experience, such as the frequency of use of particular instances of constructions, have an impact on representation that we can see evidenced in various ways...

However, much of her cited evidence conflates linguistic experience with real-world experience. For example, in the phonological reduction of extremely frequent phrases such as *I don't know*, is this reduction due to the frequency of the linguistic expression per se, or is it due to the frequency of the event of not knowing something? Similarly, in the case of Arnon and Snider's contrast between phrases such as *Don't have to worry* and *Don't have to wait*, there could be a difference in the real-world likelihood of the events described by these expressions, which causes faster processing due to the difference in conceptual predictability, as opposed to linguistic predictability.[2] In general, this confound between linguistic experience and real-world knowledge exists whenever one compares expressions describing different real-world events.

Another outstanding question is how to empirically measure the trade-off between the reuse of stored multi-word expressions and the compositional generation of expressions. In the case of novel or infrequently attested expressions, we assume that such expressions must be processed compositionally using abstract linguistic knowledge—that is, generalized knowledge that is not bound to specific lexical items or expressions. In the case of frequently attested expressions, two potential processing strategies exist: compositional generation or reuse of stored holistic representations. Previous experimental work has primarily focused on the question of whether there is *any* reuse of stored multi-word expressions, and has suggested that there is at least some, but it remains possible that even very frequent and conventionalized multi-word expressions could in part or at times also be generated anew using abstract knowledge. Thus the major question now is *to what extent* both holistic reuse and compositional generation play a role in language processing (Wiechmann, Kerz, Snider, & Jaeger,

[1] It may be possible to accommodate frequency effects for multi-word expressions under this theory, depending upon further details of the parser. In particular, processing of later words in an expression could be conditioned upon earlier words, thus creating an overall frequency difference. But this is not a direct prediction of the words-and-rules theory.

[2] Arnon and Snider did attempt to control for this real-world likelihood difference by collecting plausibility ratings for their materials, which they demonstrated did not differ in plausibility between conditions. However, plausibility in all conditions was very high, so extent differences may not have been detected due to ceiling effects.