Cognition 133 (2014) 140–155

Contents lists available at ScienceDirect

Cognition

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

Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation

Scott Seyfarth *

Department of Linguistics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0108, United States

ARTICLE INFO

Article history:
Received 18 July 2013
Received in revised form 12 June 2014
Accepted 16 June 2014

Keywords:
Probabilistic reduction
Acoustic reduction
Lexical representation
Speech production
Predictability
Informativity

ABSTRACT

Language-users reduce words in predictable contexts. Previous research indicates that reduction may be stored in lexical representation if a word is often reduced. Because representation influences production regardless of context, production should be biased by how often each word has been reduced in the speaker's prior experience. This study investigates whether speakers have a context-independent bias to reduce low-informativity words, which are usually predictable and therefore usually reduced. Content word durations were extracted from the Buckeye and Switchboard speech corpora, and analyzed for probabilistic reduction effects using a language model based on spontaneous speech in the Fisher corpus. The analysis supported the hypothesis: low-informativity words have shorter durations, even when the effects of local contextual predictability, frequency, speech rate, and several other variables are controlled for. Additional models that compared word types against only other words of the same segmental length further supported this conclusion. Words that usually appear in predictable contexts are reduced in all contexts, even those in which they are unpredictable. The result supports representational models in which reduction is stored, and where sufficiently frequent reduction biases later production. The finding provides new evidence that probabilistic reduction interacts with lexical representation.

1. Introduction

1.1. Probabilistic reduction

In speech production, language-users reduce words when they are predictable in the local context, as well as when they are frequent overall (Gahl, 2008; Lieberman, 1963; Whalen, 1991). This reduction manifests as a broad array of articulatory and acoustic effects, including differences in word and syllable duration, vowel dispersion and quality, plosive voice onset time, syllable deletion, and language-specific segmental deletion, among others (Aylett & Turk, 2006; Baker & Bradlow, 2009; Bell et al., 2003; Bybee, 2002, 2006; Clopper & Pierrehumbert, 2008; Demberg, Sayeed, Gorinski, & Engonopoulos, 2012; Everett, Miller, Nelson, Soare, & Vinson, 2011; Gahl & Garnsey, 2004; Hooper, 1976; Jurafsky, Bell, Gregory, & Raymond, 2001; Kuperman & Bresnan, 2012; Moore-Cantwell, 2013; Tily et al., 2009; Yao, 2009). These phenomena have been known for over a century (see Bell, Brenier, Gregory, Girand, & Jurafsky, 2009, for a review), and are usually described together as the *probabilistic reduction hypothesis*—words with higher probability are articulatorily reduced, for a variety of local and global probabilistic measures.

The cause of probabilistic reduction is not fully understood, although it can be accounted for in several different (and compatible) models of speech production. For example, such reduction may indicate that speakers actively manage their productions to balance audience-design

* Tel.: +1 858 534 1145.
E-mail address: sseyfarth@ucsd.edu

http://dx.doi.org/10.1016/j.cognition.2014.06.013
0010-0277/

considerations with articulatory efficiency (Lindblom, 1990). Under this theory, speakers hyper-articulate unpredictable words in order to improve listeners' chances of parsing words that they have low expectations for. They hypo-articulate words that listeners can easily predict based on the context, in order to save on articulatory effort. Smooth-signal or uniform-information-density versions of this theory frame this behavior as speakers' preference for keeping a constant rate of information transfer (Aylett & Turk, 2004; Levy & Jaeger, 2007; Pluymaekers, Ernestus, & Baayen, 2005). Speakers spend more time on unpredictable words, which are informative, and relatively little time on predictable words, which provide less new information.

An alternative account for probabilistic reduction is based in speaker-internal processing factors (Bard et al., 2000; Bell et al., 2009; Munson, 2007). Under this theory, words are activated more strongly by their phonological, semantic, and syntactic associates. This facilitates retrieval and speeds production (Gahl, Yao, & Johnson, 2012, cf. Baese-Berk & Goldrick, 2009). For example, Kahn and Arnold (2012) show that a linguistic prime causes speakers to reduce a word target even when audience-design factors are controlled for, while a non-linguistic prime does not trigger reduction.

1.2. Is reduction stored in lexical representation?

An important question is whether probabilistic reduction is exclusively an online effect, or whether it is also represented offline in the lexicon. It is generally argued that unreduced citation forms have a privileged representational status (Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004; Ranbom & Connine, 2007). However, there is evidence that reduced forms are also represented. Lavoie (2002) and Johnson (2007) show that words with homophonous citation forms can have very dissimilar distributions of reduced variants in conversational speech, and each word may in fact have special reduced variants that are unattested for its homophone. For example, [fɻ] and [fə] are attested variants of *for* but not of *four*. This suggests that reduced forms are to some extent word-specific, and therefore associated with lexical representation, rather than created exclusively online during production.

Furthermore, language-users have a processing advantage for common reduced forms of a word. This advantage is relative to how often the word is reduced (Connine & Pinnow, 2006; Connine, Ranbom, & Patterson, 2008). For example, French *genou* [ʒənu] is often realized in a reduced form [ʒnu], which lacks an audible schwa. On the other hand, *querelle* [kəʁɛl] is more often realized with a full schwa in the first syllable. In isolated word production, speakers are faster to produce forms like [ʒnu] than [kʁɛl], all else held equal, where [ʒnu] but not [kʁɛl] is a common word-specific reduction (Bürki, Ernestus, & Frauenfelder, 2010; Racine & Grosjean, 2005). In lexical decision experiments, Ranbom and Connine (2007) and Pitt, Dilley, and Tat (2011) show that listeners are faster to classify reduced forms like English *gentle* [dʒɛɾ̃ɫ], with a nasal flap, than [dʒɛnʔɫ], where the flap but not the glottal stop is a usual reduction of [t] in words like *gentle*. These findings indicate that reduced variants, when they are typical realizations of a word, are likely stored in representation (Ernestus, 2014; Pitt, 2009).

There are at least three ways this storage might be implemented. First, storage of reduction might involve multiple phonologically-abstract, categorical variants, which include both unreduced and reduced forms of a word (as described above). Second, individual productions of reduced words might be stored as exemplars with fine-grained phonetic detail, including acoustic reduction (Johnson, 2007; Pierrehumbert, 2002). Third, reduction might be represented indirectly via changes to articulatory timing relations that are lexically specified (Browman & Goldstein, 1990; Byrd, 1996; Lavoie, 2002).

Is probabilistic reduction stored in lexical representation? Reduction associated with high contextual probability is standardly treated as an online phenomenon, such as a kind of priming or else active management of information density, as in 1.1. The evidence discussed here suggests that reduction is stored when it occurs often enough. Therefore, if a word is very often reduced because it typically occurs in high-probability contexts, language-users may store this reduction in lexical representation as well.

1.3. Informativity

In usage, some words almost always occur in predictable contexts, whereas others are unlikely in each of the contexts that they occur in, even though they might be relatively frequent overall. For example, the word *current* usually occurs in the context of *current events* or *the current situation*, and is therefore usually predictable in context. On the other hand, the word *nowadays* has roughly the same log-frequency overall as *current*, but *nowadays* occurs in a wide variety of contexts (see Fig. 1). Thus, on average, *nowadays* is more unpredictable in each of its contexts.

The average predictability of a word in context is its *informativity* (Cohen Priva, 2008; Piantadosi, Tily, & Gibson, 2011). Word informativity is formally defined as:

$$-\sum_{c} P(C=c|W=w)\log P(W=w|C=c) \quad (1)$$

In Eq. (1), c is a context and w is a word type. Context is usually operationalized simply as the n preceding or following words in an utterance. The informativity of a word type is the averaged probability with which a word will occur given each of the contexts that it can occur in. This average is weighted by the frequency with which the word occurs in each context. Usually-predictable words (like *current*) have low informativity, because they tend to provide less new information in actual communicative use. Usually-unpredictable words (*nowadays*) have high informativity, because in actual use they tend to be surprising and informative.

Because low-informativity words are usually predictable, they are also usually reduced. On the other hand, high-informativity words are rarely reduced. The experiments described in 1.2 demonstrate that reduced forms of a word are more accessible if a reduced form is a typical realization of that word. If probabilistic reduction is stored, reduction of low-informativity words should be more accessible than reduction of high-informativity words. In a