Cognition 160 (2017) 27–34

Contents lists available at ScienceDirect

Cognition

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

Brief article

Not so fast: Fast speech correlates with lower lexical and structural information

Uriel Cohen Priva

Brown University, Box 1821, 190 Thayer St, Providence, RI 02912, United States

ARTICLE INFO

Article history:
Received 12 August 2015
Revised 29 August 2016
Accepted 6 December 2016
Available online 28 December 2016

Keywords:
Language
Information
Speech rate
Information rate

ABSTRACT

Speakers dynamically adjust their speech rate throughout conversations. These adjustments have been linked to cognitive and communicative limitations: for example, speakers speak words that are contextually unexpected (and thus add more information) with slower speech rates. This raises the question whether limitations of this type vary wildly across speakers or are relatively constant. The latter predicts that across speakers (or conversations), speech rate and the amount of information content are inversely correlated: on average, speakers can either provide high information content or speak quickly, but not both. Using two corpus studies replicated across two corpora, I demonstrate that indeed, fast speech correlates with the use of less informative words and syntactic structures. Thus, while there are individual differences in overall information throughput, speakers are more similar in this aspect than differences in speech rate would suggest. The results suggest that information theoretic constraints on production operate at a higher level than was observed before and affect language throughout production, not only after words and structures are chosen.

1. Introduction

1.1. Background

Do fast speakers use language differently than slower speakers? There is ample evidence that speakers slow down when upcoming material is not available (Fox Tree & Clark, 1997) or not given in the discourse (Arnold, Fagnano, & Tanenhaus, 2003). Bell, Brenier, Gregory, Girand, and Jurafsky (2009) also argue for slowdown when the word itself takes longer to retrieve. Other studies show that speakers add and remove linguistic material in response to the availability of upcoming context or the information that material provides (Ferreira & Dell, 2000; Jaeger, 2010). Given this body of evidence, we may expect that fast speakers would use language in the same way slower speakers do, and slow down when their own cognitive or communicative constraints require them to.

To account for previous findings, the cognitive and communicative constraints involved can vary substantially between individuals, and may operate at relatively low levels of language production: only at lexical retrieval, only at syntactic planning, etc. However, it is also possible that such constraints are less variable across individuals and operate at a higher level, affecting language production as a whole. In this case, speech rate and information content are expected to trade off against each other: Speakers who use more informative content would be more likely to speak more slowly, and speakers who speak fast would be more likey to use less informative content.

1.2. Information theoretic accounts

There is a growing body of research on the role of information theoretic constraints in human language. Multiple studies show that speakers (and writers, Genzel & Charniak, 2002) tend to not provide too much or too little information at any given time (Aylett & Turk, 2004; Jaeger, 2010; Levy & Jaeger, 2007), as predicted by information theory (Shannon, 1948; for a comprehensive review, see Jaeger & Buz, in press).[1] It has been proposed that speakers respond to information troughs by omitting, reducing, or hypo-articulating low-information linguistic material or to information peaks by expanding or hyper-articulating high-information linguistic material. Expansion and reduction are expected in a relatively local domain, and have been demonstrated for individual segments (Cohen Priva, 2015; van Son & Pols, 2003; van Son & van Santen, 2005), syllables (Aylett & Turk, 2004), morphemes (Kuperman,

[1] Response to information theoretic pressures can follow from both speaker-internal and communicative pressures (Jaeger, 2010, pp. 50–51 for speaker-internal alternative; and Pate & Goldwater, 2015, for communicative-based focus).

E-mail address: uriel_cohen_priva@brown.edu

http://dx.doi.org/10.1016/j.cognition.2016.12.002
0010-0277/

Pluymaekers, Ernestus, & Baayen, 2007; Kurumada & Jaeger, 2015; Pluymaekers, Ernestus, & Baayen, 2005), words (Arnon & Cohen Priva, 2014; Bell et al., 2009; Jurafsky, Bell, Gregory, & Raymond, 2001; Mahowald, Fedorenko, Piantadosi, & Gibson, 2013; Seyfarth, 2014), and notably even at the edge of clauses (Jaeger, 2010; Levy & Jaeger, 2007; Norcliffe & Jaeger, 2014), suggesting that information theoretic considerations are also driven by syntactic information. Other studies suggest that higher-level syntactic considerations affect the duration of individual words within that construction (Gahl & Garnsey, 2004; Kuperman & Bresnan, 2012).

1.3. Sources of information

Information theoretic accounts define information as surprisal: the less predictable the message, the more information it provides. Several factors affect the amount of information provided by speakers. Consider the phrase *dog bites man*. The information provided by the phrase is the negative log probability of observing the phrase, given what we know about the world, and about the English language. I consider three sources of information: *world knowledge*, *lexical*, and *structural*. World knowledge determines that dogs biting men is more probable than men biting dogs. Therefore, *dog bites man* provides less information than *man bites dog*, even though both phrases have identical words and structure. Lexical information contributes to the information a phrase provides. The word *human* is less frequent than the word *man* (despite denoting a larger set of individuals). The phrase *dog bites human* is therefore lexically more informative than *dog bites man*. Syntactic choices also contribute to the information a phrase provides. Active voice is used more frequently than passive voice. The phrase *dog bites man* is therefore structurally less informative than *man bitten by dog*, even though they relate the same event.

1.4. Speech rate and information rate

Information rate can be estimated by dividing the information provided by linguistic material with the time it takes to produce that material.[2] To keep information rate constant, speakers should slow down when providing more information and speed up when providing less. If speakers' production approximates a ratio between information and time as studies suggest, what are the implications of fast speech rate? Do fast speakers provide more information per second by approximating a higher information to time ratio? Previous studies seem to support this view, as fast speech rate is a strong predictor of the omission of linguistic material, e.g. *that*-omission in Jaeger (2010), and segment deletion in Cohen Priva (2015). A positive correlation between fast speech and omission would lead fast speakers to provide an even higher information rate than had they kept the omitted linguistic material. However, at least at the segmental level, omission could well be one of the mechanisms that make fast speech fast (e.g. by articulatory undershooting of the target).

The alternative is that fast speech rate corresponds to lower information content (evident in cross-linguistic differences, Pellegrino, Coupé, & Marsico, 2011). This possibility has surprising implications: it predicts that fast speakers may use less informative words, simpler syntactic structures, or provide less informative world knowledge, thus facilitating production and comprehension to compensate for faster speech (or vice versa). World knowledge is beyond the scope of this paper, but it is possible to test the first two predictions by investigating how speech rate correlates with lexical information and the use of infrequent syntactic structures. In the following sections I present two corpus studies that test the prediction that fast rate of speech would correlate with lower lexical and structural information rate using the Switchboard corpus (Godfrey & Holliman, 1997), and replicate them using the Buckeye corpus (Pitt et al., 2007).

[2] Other interpretations can include information over *amount* of linguistic material or cost of making a message less confusable (Jaeger & Buz, in press).

2. Studies overview: materials and methods

2.1. Averaging data across conversation sides

This study aims to investigate the relationship between different aspects of information rate *beyond* the local contexts in which they have been studied in the past. Therefore, I aggregated data from individual tokens across conversation sides rather than investigate individual tokens separately. Thus, each conversation side (one speaker's speech in one conversation) constitutes a single data point.

2.2. Corpora

I used the Switchboard Corpus (Godfrey & Holliman, 1997) to run the main studies. Each conversation provides two data points: the two sides of the conversation. I used Calhoun et al. (2009), which provides part of speech tags for a subset of the original corpus. The Buckeye corpus (Pitt et al., 2007) was used to replicate the findings from the main studies. In Buckeye only the interviewee side is available, and the 40 interviewees provide 40 data points. I retagged Buckeye using a POS-tagger (Toutanova, Klein, Manning, & Singer, 2003) for consistency with Switchboard. Words whose duration surpassed 5s were removed. To get robust estimation for word counts, word count information was pooled from the Switchboard, Buckeye and Fisher (Cieri, Graff, Kimball, Miller, & Walker, 2005, part 2) corpora. The full procedure of curating data is described in Appendix A.

2.3. Data exclusion

For both studies, utterance-final words and words that were followed by filled pauses or backchannels (e.g. *uh*, *yeah*; POS `UH`) were excluded to avoid a possible confound due to phrase-final lengthening. Only utterances 4 words or longer were used to exclude other backchannels. To reduce the possible effect of outliers, words whose log durations were not within 3 standard deviations from the mean were also removed. The exclusion criteria for each study are summarized in the methods and materials section for each study (Sections 3.2 and 4.2). Not excluding pre-pausal and utterance final words did not lead to qualitative differences, and neither did using utterances of any length.

2.4. Speech rate

In order to estimate speech rate, I had to establish how fast a word was expected to be given previous research. I therefore defined *pointwise speech rate* as the *actual duration* of a word token, divided by that token's *expected duration*. Thus, if a word's duration was predicted to be 250 ms but was pronounced in 300 ms, its pointwise speech rate would be 1.2 (slow), while if that word were pronounced in 200 ms, its pointwise speech rate would be 0.8 (fast).

Expected duration was calculated using a linear regression. The log actual duration was the predicted value, and the predictors were: (a) The geometric mean duration of that word across the corpus in which the word appeared, (b) the log probability of observing the word given the previous word, (c) the log probability of the word given the following word, (d) the log probability of the