Not so fast

Fast speech correlates with lower lexical and structural information

Uriel Cohen Priva
Brown University

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, talkers 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.

Accepted manuscript version.

Journal: *Cognition*

DOI: 10.1016/j.cognition.2016.12.002

Keywords: language; information; speech rate; information rate

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 and Clark, 1997) or not given in the discourse (Arnold et al., 2003). Bell et al. (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 and 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 possible that such constraints are less variable across individuals and operate at a higher level, affecting language production as a whole. In this case, fast speech rate should correlate with less informative content: 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 and Charniak, 2002) tend to not provide too much or too little information at any given time (Aylett and Turk, 2004; Levy and Jaeger, 2007; Jaeger, 2010), as predicted by information theory (Shannon, 1948; for a comprehensive review, see Jaeger and Buz, accepted).[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 (van Son and Pols, 2003; van Son and van Santen, 2005; Cohen Priva, 2015), syllables (Aylett and Turk, 2004), morphemes (Pluymaekers et al., 2005; Kuperman et al., 2007; Kurumada and Jaeger, 2015), words (Jurafsky et al., 2001; Bell et al., 2009; Mahowald et al., 2013; Seyfarth, 2014; Arnon and Cohen Priva, 2014), and notably even at the edge of clauses (Levy and Jaeger, 2007; Jaeger, 2010; Norcliffe and 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 and Garnsey, 2004; Kuperman and 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

[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 and Goldwater, 2015, for communicative-based focus).

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 et al., 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 and Holliman, 1997), and replicate them using the Buckeye corpus (Pitt et al., 2007).

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 *outside* 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 and 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

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

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 et al., 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 et al., 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 and 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 (§ 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 250ms but was pronounced in 300ms, its pointwise speech rate would be 1.2 (slow), while if that word were pronounced in 200ms, 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 previous word given the current word, (e) the log probability of the following word given the previous word, and (f) the frequency of the three words together. Word counts for variables (b–f) were taken from the Fisher, Switchboard and Buckeye corpora.[3] The controls were meant to account for previously-studied factors, including the phonological form of the word, frequency effects (e.g. Zipf, 1935; Bybee, 2002), the availability of the following word (e.g. Fox Tree and Clark, 1997), transitional probabilities (e.g. Seyfarth, 2014) and trigram frequency (Arnon and Cohen Priva, 2013, 2014). This process is exemplified in Table 1.

Table 1: An example of the predictors used to approximate expected duration, compared with actual duration. The linear regression uses all the information theoretic variables as well as the log mean word duration to predict log word duration. Word duration is then divided by the exponentiated predicted value to yield pointwise speech rate for each word.

Variable	Example 1	Example 2
word duration	0.16	0.23

[3] Variables (b–e) were calculated using MLE, as the count of observing the word sequence, divided by the count of the given context alone. The previous word in utterance-initial contexts was taken to be a special "missing" context, as was the following word in utterance-final contexts.

Variable	Example 1	Example 2
previous word	start	and
word	going	tell
following word	to	them
mean word duration	0.203	0.220
-logPr(word | previous)	6.03	10.50
-logPr(word | next)	4.04	6.40
-logPr(previous | word)	8.66	4.57
-logPr(next | word)	0.784	4.183
trigram frequency	20	35
pointwise speech rate	0.923	1.130

Each speaker's average by-conversation *speech rate* was defined as the average of log pointwise speech rates in that conversation side. Calculating speakers' speech rate using medians changed the significance level of speech rate in the first study using the Switchboard corpus to marginal (p=0.058).

2.5 Information Rate

In this paper I use two criteria for information rate: lexical information rate and structural information rate.

2.5.1 Lexical information rate

The lexical information for each word was calculated as the negative probability of observing that word using a unigram model with combined counts from the Buckeye, Switchboard, and Fisher corpora. Each speaker's average by-conversation *lexical information rate* was defined as the average of the negative log probabilities of observing all content words used to calculate speech rate in that conversation side. This is the *entropy* of each speaker's content words. Content words were defined as all words which were not filled pauses, backchannels (both POS UH), or function words (words that signal grammatical relations; defined as the output of the function `stopwords()` in the R package `tm`, Feinerer and Hornik, 2015, as well as the contracted forms of *be* and *have*). Function words were not used as they have been argued to be retrieved using different mechanisms than content words (Bell et al., 2009).

2.5.2 Structural information rate

Because the Buckeye Corpus is not parsed and automated parsing is not as accurate as automated part-of-speech tagging, I chose a syntactic structure that could be relatively easily detected using part of speech tags for this study. I chose *passive voice*, a structure that is *optional* in the sense that speakers can, for the most part, choose the equivalent active voice without omitting information.

The vast majority of sentences in spoken corpora are in the active voice as in (1), rather than in passive voice (2) or (3). This means that the probability of observing an active voice sentence is higher than the probability of observing a passive voice sentence, everything else being equal. Speakers who avoid passive voice in favor of the alternative, active voice, provide more probable and therefore less informative structures.

(1) The police arrested John.

(2) John was arrested (by the police).

(3) John got arrested (by the police).

The prediction is that slow speech would be correlated with the use of the informative and infrequent passive voice (or vice versa).

I defined passive voice in terms of part of speech transitions: some inflection of *be* or *get*, followed by a past participle verb (POS VBN). This definition excludes cases in which an adverb or negation intervene between the auxiliary verb and the past participle. Since the goal of this paper is not to count passive voice sentences, this choice should not be problematic.

Speaker's by-conversation log *rate of passives* was calculated as the log odds between passive voice clauses, as defined above, and other types of clauses, counted using the number of active voice content verbs in that conversation side. Log odds were used in order to allow parameter estimation using a linear regression.

3 Study 1: Speech rate and lexical information

3.1 Introduction

If speakers maintain a relatively constant information rate, they will compensate for speaking faster by providing less lexical information, or vice versa. This study checks whether fast speakers do use more frequent words (estimated by a unigram model).

3.2 Methods and materials

I fitted a mixed effects linear regression in R (R Core Team, 2015) using the lmerTest package (Kuznetsova et al., 2014), which fits models using the lme4 package (Bates et al., 2014) and adds p-values. For the replication study in Buckeye there are no repeated measures, and so a linear regression was used instead. The predicted variable was speaker's average by-conversation lexical information rate, as discussed in § 2.5.1. Speaker's average by-conversation speech rate (normalized), as discussed in § 2.4, was the variable of interest. Both lexical information and speech rate were calculated using only the words in Table 2.

Table 2: Data exclusion summary for Study 1

Criterion	Switchboard words	Switchboard % removed	Buckeye words	Buckeye % removed
All content words in the corpus	295627	0	115220	0
After exclusion of word longer than 5s	-	-	114660	0.49
After removal of utterance final words, and pre-pausal words	265289	10.26	90999	20.64
After removal short utterances (under 4)	264185	0.42	85759	5.76
After contextual counts were added	253080	4.2	85458	0.35

Criterion	Switchboard words	Switchboard % removed	Buckeye words	Buckeye % removed
After removal of extreme durations	253047	0.01	85355	0.12

For the Switchboard data, the speaker's gender, log age, and the interlocutor's speech rate (normalized) were used as controls. Expected topic lexical information rate (the average lexical information rate in the topic across speakers) was also controlled for, as some topics may correlate with less probable words. Speakers and topics were used as random intercepts. The model was fitted using 1284 conversation sides, which included 358 speakers and 64 topics.

For the replication in Buckeye, interviewer identity, gender, and Buckeye's binary distinction for age were used as controls. All 40 speakers constitute the conversation sides in the replication study.

The Switchboard and Buckeye controls are summarized in Tables 3 and 4.

Table 3: Controls used for Study 1, in Switchboard

Control	Description	Motivation
Speaker's gender	Binary	Male and female speakers have demonstrated different speech patterns
Age	The log of the speaker's age	Age may effect words speaker uses
Interlocutor's speech rate	Mean log pointwise speech rate of interlocutor (normalized)	Interlocutor speech rate may affect partner's speech through alignment
Expected topic lexical information rate	Mean negative log probability of all content words over conversations on that topic	Some topics may correlate with less probable words

Table 4: Controls used for Study 1, in Buckeye

Control	Description	Motivation
Interviewer ID	Binary	Speakers may have been affected by interviewer
Speaker's gender	Binary	Male and female speakers have demonstrated different speech patterns
Age	Binary	Numeric age not recorded by Buckeye. Age may effect words speaker uses

3.3 Results and discussion

As predicted, slow speech rate correlated with higher average lexical information (β=0.055, SE=0.02, t=2.704, p<0.01). Expected topic lexical information rate significantly predicted speakers' unigram lexical information rate (β=0.23, SE=0.013, t=16.912, $p<10^{-15}$). Male speakers and older speakers were more likely to use infrequent words (β=0.31, SE=0.048, t=6.386, $p<10^{-9}$; β=0.35, SE=0.083, t=4.175, $p<10^{-4}$ respectively). Interlocutor speech rate (β=-0.018, SE=0.013, t=-1.333, p=0.183) did not affect speakers' lexical information rate.

Figure 1 shows by gender the relationship between normalized speech rate and average unigram lexical rate in Switchboard. The lines represent the raw correlation between speech rate and lexical information rate by gender, and do not factor out other predictors. Each point in the figure represents a conversation side, marked for gender. All following figures repeat this scheme.

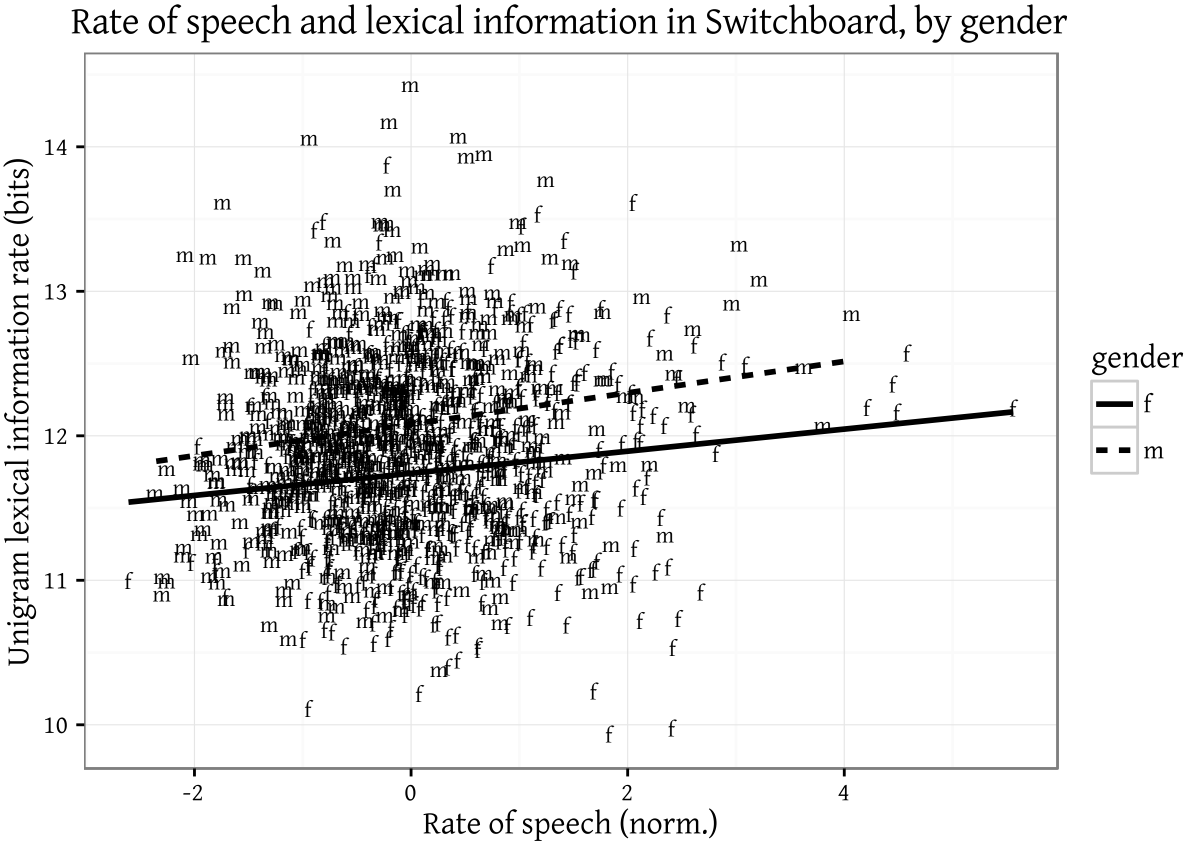

Figure 1: Positive correlation between slow speech rate, measured in mean pointwise speech rate and normalized, and lexical information rate, measured in mean negative log probability (bits) in Switchboard. Each point represents a speaker in a conversation. The lines represent the raw correlations between speech rate and unigram lexical information rate, by gender.

Buckeye corpus results support the findings in Switchboard: slow speech rate correlated with higher lexical information rate (β=0.16, SE=0.071, t=2.313, p<0.05). Male speakers were again more likely to use infrequent words (β=0.33, SE=0.14, t=2.375, p<0.05). Age (β=-0.21, SE=0.14, t=-1.508, p=0.141) and interviewer identity (β=-0.14, SE=0.14, t=-1.054, p=0.299) did not affect lexical information rate.[4] Figure 2 shows by gender the correlation between average speech rate and average unigram lexical information.

The results suggest that unigram lexical information and speech rate come at the expense of one another. Fast speech correlated with less informative words. The gender findings are surprising, and may follow from so-

[4] In order to verify that the results do not depend on outliers, I reran the regression on subsets of that data in which one speaker was removed. There were three speakers in the Buckeye corpus such that the results became marginal if any one of them were removed (p<0.067).

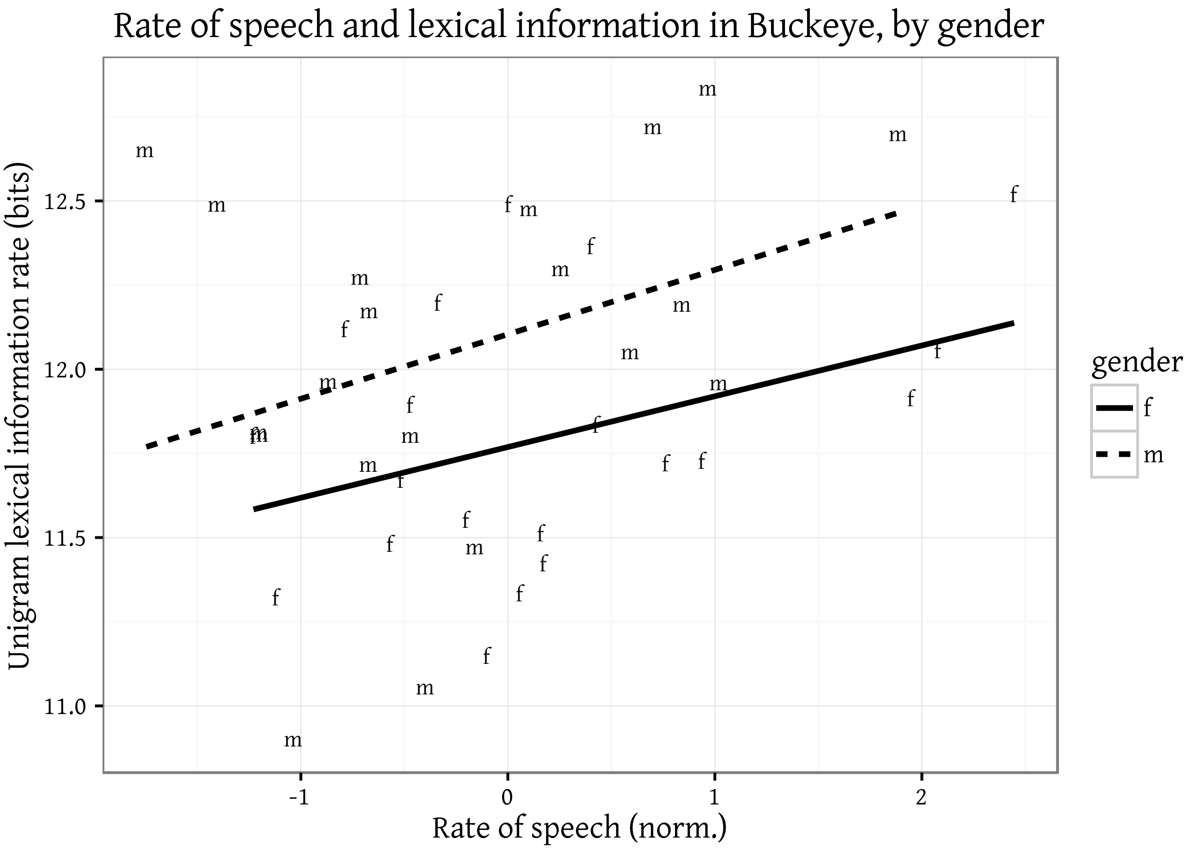

Figure 2: Positive correlation between slow speech rate, measured in mean pointwise speech rate and normalized, and lexical information rate, measured in mean negative log probability (bits) in Buckeye. Each point represents a speaker. The lines represent the raw correlations between speech rate and unigram lexical information rate, by gender.

ciolinguistic factors.[5] The correlation between older age and use of infrequent words may follow from older people's familiarity with outdated lexical items, a hypothesis this study does not test.

4 Study 2: speech rate and passive voice

4.1 Introduction

If speakers maintain a relatively constant information rate, they will compensate for speaking faster by providing less informative structures, or vice versa. This study checks whether fast speakers are more likely to use active voice over passive voice.

4.2 Methods and materials

As in study 1, I used a mixed effects linear regression to model data from Switchboard, with speaker and topic as random intercepts, and a linear regression for Buckeye. For both corpora I used speaker's average by-conversation speech rate as the variable of interest. Speaker's by-conversation log rate of passives was the predicted variable. Controls are the same as in study 1 (see Tables 3 and 4), except for by-topic rates in Switchboard, which encode average rate of passive voice for the topic of conversation, rather than average lexical information rate.

Since passive voice is an infrequent construction, many speakers never used passive voice in a conversation. Such data points were excluded from the analysis (or log ratio would have been undefined), resulting in 783 data points for Switchboard (61% of all available conversation sides). All speakers in Buckeye used passive voice at least once. Speech rate in this study was calculated using only the active-voice verbs. For summary of exclusions see Table 5.

Table 5: Data exclusion summary for Study 2

Criterion	Switchboard words	Switchboard % removed	Buckeye words	Buckeye % removed
All verbs in the corpus	133739	0	52838	0
After exclusion of words longer than 5s	-	-	52701	0.26
After removal of utterance final verbs, and pre-pausal words	125150	6.42	45017	14.58
After removal short utterances (under 4)	124521	0.5	42645	5.27
After contextual counts were added	121517	2.41	42560	0.2
After removal of extreme durations	121497	0.02	42507	0.12
Number of active content verbs	81152	-	41812	-

[5] Models that controlled for education were comparable. In one model college education yielded higher lexical information.

Criterion	Switchboard words	Switchboard % removed	Buckeye words	Buckeye % removed
Total number of passive voice verbs	1783	-	695	-

4.3 Results and discussion

As predicted, slow speech rate was correlated with frequent use of passive voice (β=0.078, SE=0.024, t=3.217, p<0.01). Different topics had typical use odds for passive voice (β=0.24, SE=0.021, t=11.553, p<10^{-15}). Male speakers were more likely to use passive voice (β=0.1, SE=0.049, t=2.121, p<0.05). Speaker age (β=-0.084, SE=0.087, t=-0.97, p=0.332) and the interlocutor's speech rate (β=0.013, SE=0.021, t=0.65, p=0.519) had no significant effect. Figure 3 shows the correlation between average speech rate and passive voice usage.

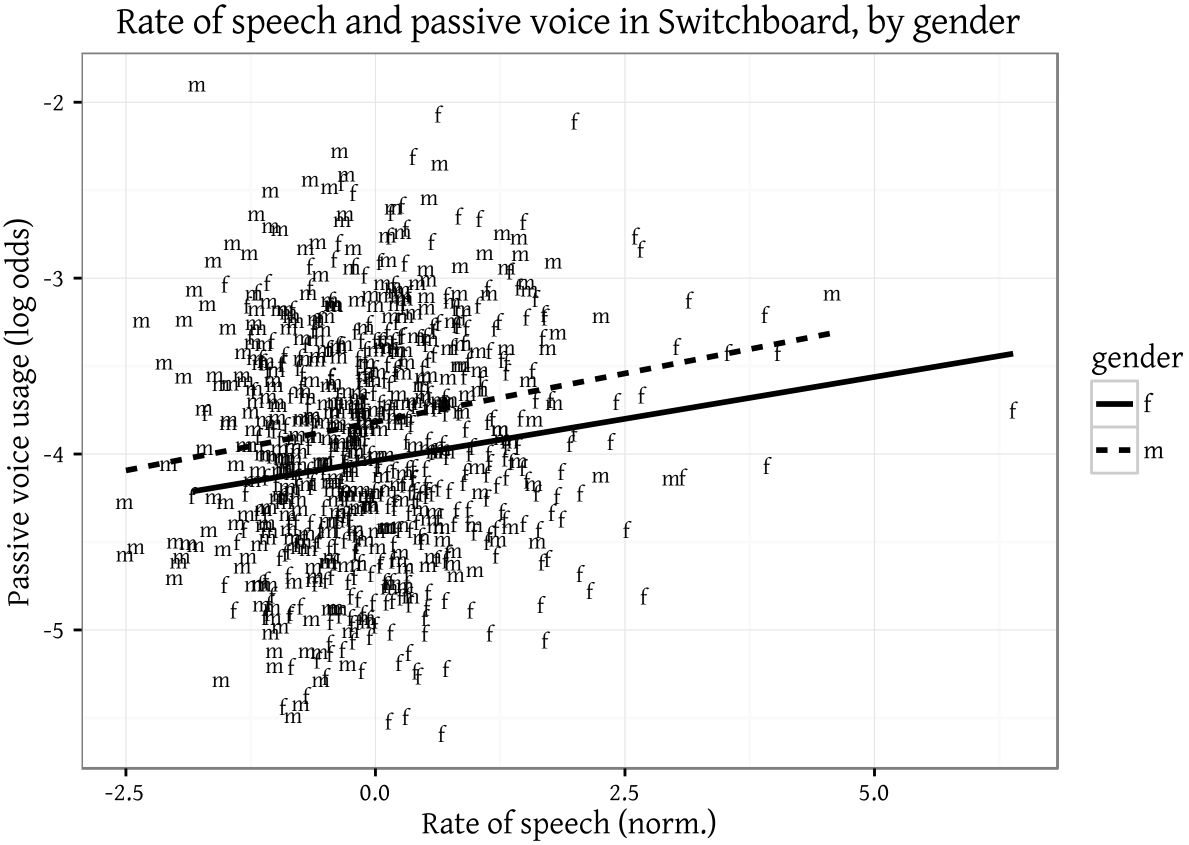

Figure 3: Positive correlation between slow speech rate, measured in mean pointwise speech rate and normalized, and passive voice usage, measured in log odds in Switchboard. Each point represents a speaker in a conversation. The lines represent the raw correlations between speech rate and passive voice usage, by gender.

Buckeye data corroborates the main finding: slow speech rate was correlated with high usage of passive voice constructions (β=0.15, SE=0.058, t=2.499, p<0.05). Gender (β=0.092, SE=0.11, t=0.8, p=0.428), age (β=-0.1, SE=0.11, t=-0.92, p=0.365) and interviewer identity (β=-0.11, SE=0.11, t=-1.006, p=0.321) did not affect the use of passive voice. The results do not depend on the inclusion of any one speaker. Figure 4 shows the correlation between average speech rate and passive voice usage.

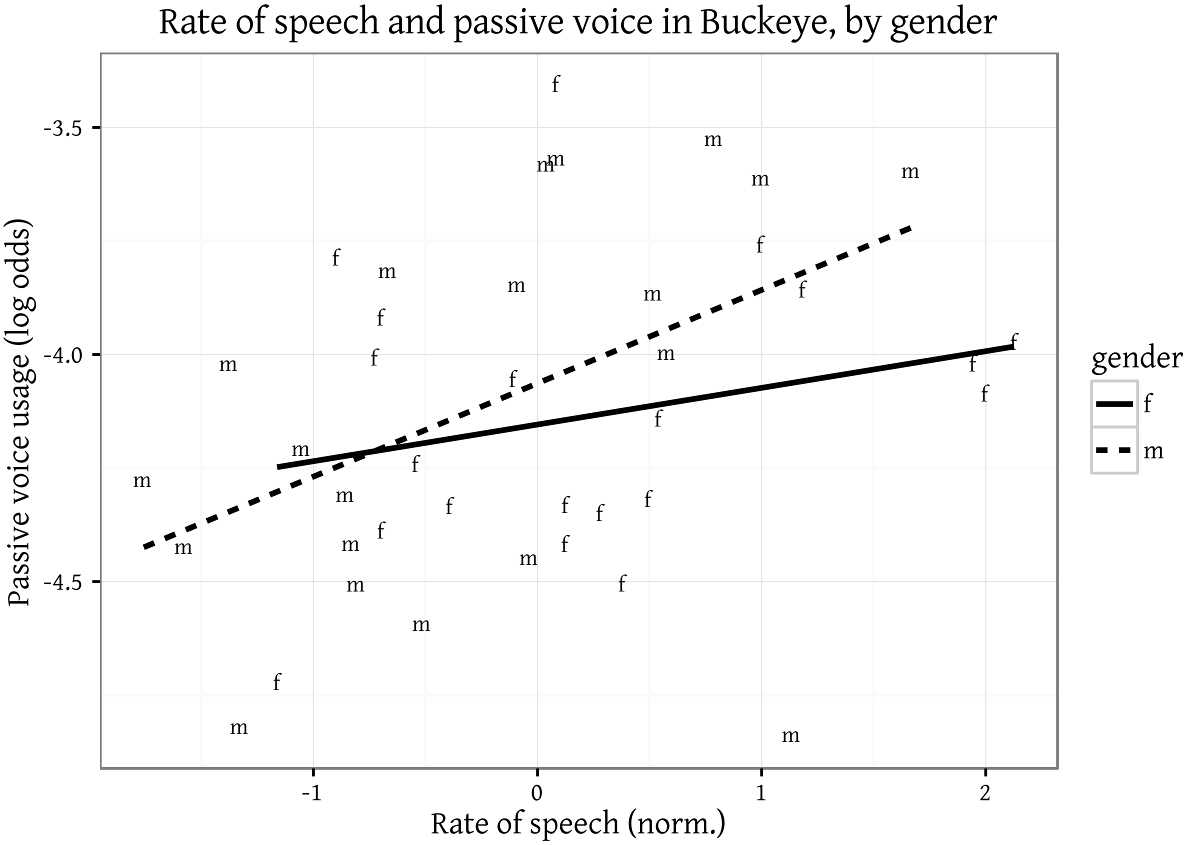

Figure 4: Positive correlation between slow speech rate, measured in mean pointwise speech rate and normalized, and passive voice usage, measured in log odds in Buckeye. Each point represents a speaker. The lines represent the raw correlations between speech rate and passive voice usage, by gender (though the difference between the two is not significant, and the slopes are not significantly different).

Study 2 provides evidence that syntactic information and speech rate come at the expense of one another. Faster speech correlates with lower usage of the structurally informative passive voice. The similar finding for gender is intriguing, but not replicated in Buckeye. As with study 1, the replication in Buckeye shows that the effect is not corpus-specific.

5 General discussion

Previous studies showed a correlation between speech rate and information rate: when speakers produce informative linguistic content, they slow down. This study suggests the effect is not only local, but applies globally. Fast speakers are likely to produce less informative content, even after previously studied local effects have been controlled for. Both lexical information and structural information are likely to be lower for fast speakers. These findings have implications on the scope of information theoretic constraints and how early in production they come into effect.

It is intriguing that the Switchboard studies could be replicated in Buckeye, with just 40 speakers. This suggests that variability in channel capacity is surprisingly small. If different people had widely different channel capacities, we would have expected many speakers who speak fast *and* provide highly informative words and structures, and other speakers who both speak slowly *and* provide less informative words and structures, making the correlation between speech rate and other kinds of information rate more difficult to find.

One possibility is that fast speakers limit their information content to speak quickly. Repeatedly choosing more frequent words and structures would mean that speakers would rarely have to slow down for infrequent words and structures, thereby maintaining overall faster speech rate. Thus the global relationship between information and speech rate is achieved, rather than only the previously found local relationship.

Being less likely to use less informative words and structures may have global benefits in production that would allow speakers to speak faster generally: Speakers who are less likely to use informative words and structures may retrieve uninformative words and frequent structures more easily, as those will not be in competition with the more informative words and infrequent structures. For speakers more likely to use infrequent words, those words get more probability mass, making frequent words less probable (in information theoretic terms, more contentful and difficult to process) than for speakers who are apriori less likely to use infrequent words. In this interpretation, being less likely to use flowery speech leads to faster speech rate by facilitating the retrieval of frequent words. Of course, the very tendency to use more or less flowery speech could be motivated by communication-oriented or sociological (and thus communicative) goals.

Another possibility is that speakers are not fully flexible, and are unable to slow down sufficiently when they provide more informative content. If that is the case, production-oriented reasoning would directly predict that fast speakers would have to avoid informative words and structures so as not to exceed the channel capacity, which may differ among individuals.

A comprehension-oriented interpretation of this possibility would be that if fast speakers do not always slow down when necessary they will inevitably provide too much information for their listeners at certain times, exceeding channel capacity. Thus, the lack of flexibility would indirectly lead speakers to opt for less informative words and structures. In this interpretation, exceeding the listener's processing capacity can be a sociolinguistic choice. This alternative gains support from the higher lexical information rate provided by male speakers in both corpora, in light of the absence of any studies pointing to an advantage for men in language production.

The findings also bear on a separate question: where do information theoretic constraints apply? In a *late model* speakers freely select any word or syntactic structure, and will later add enough linguistic material or slow down to obey information constraints. In contrast, the *throughout model* integrates information theoretic constraints even at the stage in which speakers choose words and syntactic structures (as in Jaeger, 2010, pp. 50-51). Previous findings on information theory in language could arise from either model, but the studies presented above support the *throughout model*.

Different types of information rate come at the expense of one another. Speakers do not or cannot increase an aspect of information rate without reducing another aspect. This paper provides high-level evidence for information theoretic constraints on language: not just at the segment, morpheme, word or even phrase level, but at the very selection of which words and structures one uses.

6 Acknowledgements

This paper greatly benefited from discussions with Dan Jurafsky, Emily Gleason, Fiery Cushman, Elinor Amit, Scott AnderBois, Sheila Blumstein, Meghan Sumner, and Christopher Mescher.

A Corpora

A.1 Switchboard Corpus

The Switchboard Corpus (Godfrey and Holliman, 1997) contains about 2400 annotated telephone conversations between strangers. Each speaker was paired randomly by computer operator with various other speakers; For each conversation one of 70 possible topics was assigned for discussion between speakers. Each conversation in Switchboard provides two data points: the speech of each of the conversants. Speakers and topics were used as random intercepts. Over the years many annotations were added to subsets of the corpus, adding phonetic, syntactic, and semantic information. I use a prominent collection of such annotations, Switchboard in NXT (Calhoun et al., 2009). I focused on two sources of information: duration and syntactic information, derived from part of speech (POS).

A.2 The Buckeye Corpus

The Buckeye Corpus of conversational speech (Pitt et al., 2007) contains interviews with 40 residents of Columbus OH. The smaller Buckeye corpus is used to replicate the Switchboard studies. Each interview in Buckeye provides one data point: the speech of the interviewee. Two interviewers are modeled as a binary fixed effect.

The corpus contains POS information for words, but its tags differ from Switchboard's conventions, and many backchannels such as *okay* are tagged as nouns and adjectives. I therefore trained the Stanford POS tagger (Toutanova et al., 2003) on the NXT Switchboard tags and used it to retag the Buckeye corpus.[6] The result allowed me to use POS tags that were equivalent to the ones provided by the Switchboard corpus. Buckeye includes words whose duration exceeds 5 seconds (the longest word in Switchboard is 4.5s long), and such words were removed prior to any further processing.[7] Buckeye includes the gender (M/F) and age of the interviewee (binary: younger/older than 40).

A.3 Additional corpora

Several of the controls used in the two studies required a robust estimation of word probability and word-to-word transitional probability. To achieve that, I pooled counts from the Fisher (Cieri et al., 2005, part 2), Buckeye, and Switchboard corpora. The text files used to create the count data differ minimally from the word duration files for both corpora, mostly in the treatment of interrupted speech and false starts (e.g. *y- you know*). This led to the loss of about 3% of the data in Switchboard, and 0.3% of the data in Buckeye. The loss of this data is not expected to bias the results, especially because utterance-final words and words that were followed by filled pauses were also excluded.

References

Arnold, Jennifer E., Fagnano, Maria., and Tanenhaus, Michael K. 2003. Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, 32(1):25–36.

[6] The original POS tags classified most *okay* instances as nouns, and most *right* instances as adverbs, while the new classified most instances of both as backchannels. Using the original POS tags does not have a qualitative effect on the results.

[7] A sample of several of the words that were over 5 seconds long showed that all were caused by transcription errors.

Arnon, Inbal and Cohen Priva, Uriel. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*. doi: 10.1177/0023830913484891.

Arnon, Inbal and Cohen Priva, Uriel. 2014. Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, 9(3):377–400. doi: 10.1075/ml.9.3.01arn.

Aylett, Matthew and Turk, Alice. 2004. The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.

Bates, Douglas, Maechler, Martin, and Bolker, Ben. 2014. `lme4`*: Linear mixed-effects models using S4 classes*. URL http://cran.r-project.org/package=lme4. R package version 1.1-7.

Bell, Alan, Brenier, Jason, Gregory, Michelle, Girand, Cynthia, and Jurafsky, Daniel. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1): 92–111.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14(03):261–290. ISSN 1469-8021. doi: 10.1017/S0954394502143018.

Calhoun, Sasha, Carletta, Jean, Jurafsky, Daniel, Nissim, Malvina, Ostendorf, Mari, and Zaenen, Annie, 2009. Nxt switchboard annotations. Linguistic Data Consortium Corpus. URL http://catalog.ldc.upenn.edu/LDC2009T26.

Cieri, Christopher, Graff, David, Kimball, Owen, Miller, Dave, and Walker, Kevin, 2005. Fisher English training part 2, transcripts. Linguistic Data Consortium, Philadelphia.

Cohen Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, 6(2): 243–278. URL http://www.degruyter.com/view/j/labphon.2015.6.issue-2/lp-2015-0008/lp-2015-0008.xml.

Feinerer, Ingo and Hornik, Kurt. 2015. *tm: Text Mining Package*. URL http://CRAN.R-project.org/package=tm. R package version 0.6-2.

Ferreira, Victor S. and Dell, Gary S. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4):296–340.

Fox Tree, J. E. and Clark, H. H. 1997. Pronouncing *the* as *thee* to signal problems in speaking. *Cognition*, 62:151–167.

Gahl, Susanne and Garnsey, S. M. 2004. Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language*, 80(4):748–775.

Genzel, Dmitriy and Charniak, Eugene. 2002. Entropy rate constancy in text. In *Proceedings of the Association for Computational Linguistics*, pages 199–206.

Godfrey, John J. and Holliman, Edward, 1997. Switchboard-1 release 2. Linguistic Data Consortium, Philadelphia.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

Jaeger, T. Florian and Buz, Esteban. accepted. Signal reduction and linguistic encoding. In M., Fernández Eva and Cairns, Helen Smith, editors, *Handbook of Psycholinguistics*. Wiley-Blackwell.

Jurafsky, Daniel, Bell, Alan, Gregory, Michelle L., and Raymond, William D. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, Joan L. and Hopper, Paul, editors, *Frequency and the Emergence of Linguistic Structure*, pages 229–254. Benjamins, Amsterdam.

Kuperman, Victor and Bresnan, Joan. 2012. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, 66(4):588–611. doi: 10.1016/j.jml.2012.04.003.

Kuperman, Victor, Pluymaekers, Mark, Ernestus, Mirjam, and Baayen, Harald. 2007. Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, 121 (4):2261–2271. doi: 10.1121/1.2537393.

Kurumada, Chigusa and Jaeger, T. Florian. 2015. Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, 83:152–178. doi: 10.1016/j.jml.2015.03.003.

Kuznetsova, Alexandra, Bruun Brockhoff, Per, and Haubo Bojesen Christensen, Rune. 2014. *lmerTest: Tests in Linear Mixed Effects Models*. URL http://CRAN.R-project.org/package=lmerTest. R package version 2.0-20.

Levy, Roger and Jaeger, T. Florian. 2007. Speakers optimize information density through syntactic reduction. In Scholkopf, Bernhard, Platt, John, and Hofmann, Thomas, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 19, pages 849–856, Cambridge, MA. MIT Press.

Mahowald, Kyle, Fedorenko, Evelina, Piantadosi, Steven T., and Gibson, Edward. 2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313–318. doi: 10.1016/j.cognition.2012.09.010.

Norcliffe, Elisabeth and Jaeger, T. Florian. 2014. Predicting head-marking variability in yucatec maya relative clause production. *Language and Cognition*, FirstView:1–39. doi: 10.1017/langcog.2014.39.

Pate, John K. and Goldwater, Sharon. 2015. Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, 78:1–17. doi: http://dx.doi.org/10.1016/j.jml.2014.10.003.

Pellegrino, François, Coupé, Christophe, and Marsico, Egidio. 2011. Across-language perspective on speech information rate. *Language*, 87(3):539–558. doi: 10.1353/lan.2011.0057.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E., 2007. Buckeye corpus of conversational speech (2nd release). Department of Psychology, Ohio State University.

Pluymaekers, Mark, Ernestus, Mirjam, and Baayen, R. Harald. 2005. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62:146–159.

R Core Team, . 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1):140–155. doi: 10.1016/j.cognition.2014.06.013.

Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27: 379–423.

Toutanova, Kristina, Klein, Dan, Manning, Christopher D., and Singer, Yoram. 2003. Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL ’03, pages 173–180, Stroudsburg, PA, USA. Association for Computational Linguistics. doi: 10.3115/1073445.1073478.

Son, R. J. J. H. van and Pols, L. C. W. 2003. How efficient is speech? *Proceedings of the Institute of Phonetic Sciences*, 25:171–184.

Son, R.J.J.H. van and Santen, J.P.H van. 2005. Duration and spectral balance of intervocalic consonants: a case for efficient communication. *Speech Communication*, 47:100–123.

Zipf, George Kingsley. 1935. *The Psycho-biology of Language: an Introduction to Dynamic Philology*. Houghton, Mifflin.