(In)dependence of lexical and syntactic production: *that*-reduction and omission in spontaneous speech

T. Florian Jaeger, Neal Snider, Laura Staum, and Dan Jurafsky (Stanford University)
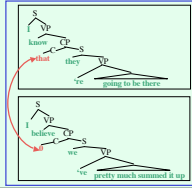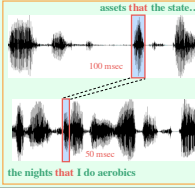**Contact:** tiflo@stanford.edu (Florian)

The Question

What is the role of probability and production difficulty in human language production? The case of reduction.

Syntactic reduction: *that*-omission

Phonetic reduction: *that*-reduction

assets **that** the state…
100 msec

the nights that **I** do aerobics
50 msec

- Recent (but separate) work on these reduction phenomena suggests that they are both probabilistic:
  - *that* is phonetically reduced when predictable (Jurafsky et al., 2001, Bell et al., 2003)
  - *that* produced less frequently if the relative/complement clause is more predictable (Wasow et al., in press, Jaeger et al., 2005, & Jaeger and Levy in progress)
- Similarly, both syntactic and phonetic reduction are affected by production difficulty (Bell et al., 2003 and Jaeger, 2005)
- → Perhaps omission (syntactic reduction) is just an extreme form of phonetic reduction (see Labov, 1969)?

Do syntactic and phonetic reduction differ in how they are affected by:
- probabilistic factors?
- experienced and/or anticipated production difficulty?
- phonological factors?

The Experiment

Methods

- Collected examples from the 800,000 word syntactically annotated Switchboard corpus of American English telephone conversations.
- We extracted **6,648 complement clauses** and **3,465 relative clauses** from the corpus.
- **1,265 complement clauses (19%)** and **1,410 relative clauses (40%)** contained complementizer or relativizer *that*.
- **Durations were automatically extracted from time-aligned orthographic transcriptions** (Stanford-Edinburgh Paraphrase Link Project).
- Time aligned transcripts at a 10msec resolution (Deshmukh, et al., 1998).
- Earlier work on phonetic reduction averaged over different uses (lemmas) of *that*. (the ICSI Switchboard of ~40,000 words; Bell et al., 2002)
- **Potentially different lemmas (complementizer and relativizer *that*) are investigated separately.**

Factors

Disfluency factors

Edit Phrase · Repetition
(1) "Well I guess (you know) I… I guess I can see it from several different perspectives."

Silence · Filled Pause
(2) "I think <silence=.34msecs> that (uh) this happens to be uh uh…"

Phonological factors
- **Preceding/Following Phone**: A following consonant increases reduction (Jurafsky et al. 1998, Bell et al. 2003)
- **Preceding/Following Stress**: A following stressed syllable increases reduction (Bell et al. 2003).

Preceding Phone · Following Phone
(3) "So I think that potentially that's affected it to some degree."
Preceding stress · Following syllable unstressed

Predictability factors

Conditional probability of RC/CC given the preceding word:
… [PP about it], …
… [NP it] through, …
… etc. …
(4) If you **think** $P[CC | think] = 0.15$ [CC (that) this is all], you are mistaken.

Conditional probability of RC/CC given its first word:
(5) $P[RC | Europe]$
That's *something* [RC (that) Europe didn't do _ right away]

Raw probability of the preceding word:
(6) …a recipe [(that) I really like] for chicken enchiladas. $P[I]$
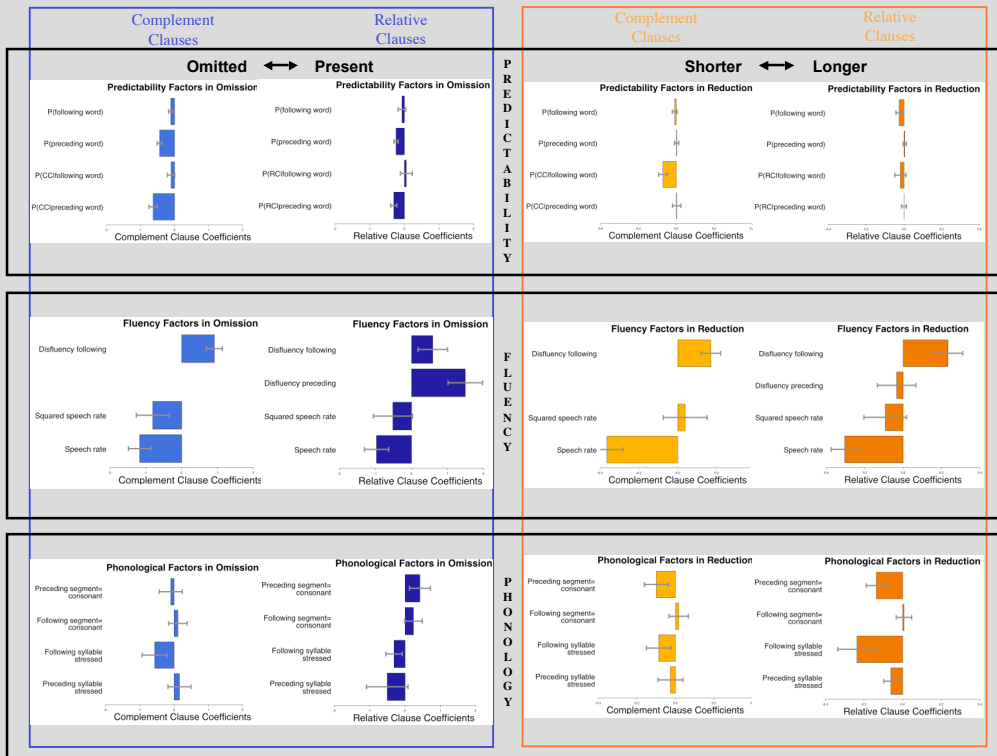
Raw probability of the following word:
(7) $P[think]$
I think [(that) the gang violence is scaring everyone to death too].

Structural predictability

Raw probabilities

Results/ Conclusions

Omission of *that*

Complement Clauses — Omitted ↔ Present
- Predictability Factors in Omission
- Fluency Factors in Omission
- Phonological Factors in Omission

Relative Clauses
- Predictability Factors in Omission
- Fluency Factors in Omission
- Phonological Factors in Omission

Reduction of *that*

Complement Clauses — Shorter ↔ Longer
- Predictability Factors in Reduction
- Fluency Factors in Reduction
- Phonological Factors in Reduction

Relative Clauses
- Predictability Factors in Reduction
- Fluency Factors in Reduction
- Phonological Factors in Reduction

PREDICTABILITY · FLUENCY · PHONOLOGY

Results: Comparison

	RC		CC	
	Omitted/ Shorter	Present/ Longer	Omitted/ Shorter	Present/ Longer
P(XC|preceding word)	***		***	
P(XC|following word)			***	
P(preceding word)	**		***	
P(following word)	**		***	
Speech rate	*** ***		*** ***	
Squared speech rate			***	
Disfluency preceding		***		
Disfluency following		** ***		***
Following syllable stressed	**		** **	
Preceding syllable stressed	* ***			
Following segment= consonant				
Preceding segment= consonant	***		***	
	Omitted/ Shorter	Present/ Longer	Omitted/ Shorter	Present/ Longer

* = significance in omission model

* = significance in reduction model

Conclusions

- **Syntactic reduction** is largely sensitive to probabilities given **preceding items**, while **phonetic reduction** is primarily influenced by **following items**
- **Adjacent disfluency** affects both phenomena in the **same way**: in disfluent contexts, *that* is more likely to be produced more frequently and longer
- Stress affects phonetic reduction and syntactic reduction in the same way. Segmental context doesn't

➔ **Omission and reduction are similar (and may partly be driven by similar mechanisms)**
➔ **Omission is NOT an extreme form of reduction**

References

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113 (2), 1001-1024.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, and Daniel Gildea. 1999. Forms of English function words - Effects of disfluencies, turn position, age and sex, and predictability. *Proceedings of ICPhS-99*.

Deshmukh, N., A. Ganapathiraju, A. Gleeson, J. Hamaker, and J. Picone. 1998. Resegmentation of Switchboard. In R. H. Mannell and J. Robert-Ribes, editors, *Proc. ICSLP*, pp. 1543–1546, Sydney.

Jaeger, T.F. 2006. Optional that-Omission: Syntactic variation or Allomorphy? Berkeley Linguistics Society 32.

Jaeger, T.F. 2005. Optional that indicates production difficulty: Evidence from disfluencies. *Proceedings of DiSS'05,* Disfluency in Spontaneous Speech Workshop. 10–12 September 2005, Aix-en-Provence, France, pp. 103-109.

Jaeger, T.F. & T. Wasow. In press. Processing as a Source of Accessibility Effects on Variation. *Proceedings of the 31st BLS*.

Jaeger, T.F., R. Levy, T. Wasow, and D. Orr. 2005. The Absence of "that" is Predictable if a Relative Clause is Predictable. Architecture and Mechanisms for Language Processing 2005, Ghent.

Jurafsky, Daniel, Alan Bell, Eric Fosler-Lussier, Cynthia Girand, and William Raymond. 1998. Reduction of English function words in Switchboard. *Proceedings of ICSLP-98*, Volume 7, 3111-3114.

Jurafsky, Daniel, Alan Bell, and Cynthia Girand. 2002. The Role of the Lemma in Form Variation. In Gussenhoven, Carlos and Natasha Warner (eds.), *Papers in Laboratory Phonology VII.* Berlin/New York: Mouton de Gruyter, 1-34.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond. 2001. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In Bybee, Joan and Paul Hopper (eds.). *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins. 229-254.

Labov, William. 1969. "Contraction, Deletion and Inherent Variability of the English Copula." Language 45: 715-62.

Wasow, T., T.F. Jaeger and D. Orr, to appear. Lexical Variation in Relativizer Frequency. In H. Simon & H. Wiese (eds.): Proceedings of the 2005 DGfS workshop "Expecting the unexpected: Exceptions in Grammar".

For further references on *that*:

www.stanford.edu/~lstaum/thatbib.htm