Instructions:
Fill in this form.
Save it with the 1st author's Last name as the file name.
If you have several submissions, number them accordingly, e.g., Bertram1.doc.
Send the form as an email attachment to amlap2007@utu.fi.
The subject line of your email should be "submission".

13th Annual Conference on Architectures and Mechanisms for Language Processing

24-27 August 2007
Turku, Finland

1. Presentation preference: ☒ Oral presentation ☐ Poster ☐ No preference

2. Topic of the presentation (please select up to 3 options or define in "other"):

☐ Word recognition
☐ Lexical processing and representation
☒ Morphology
☐ Parsing and interpretation
☐ Sentence comprehension
☐ Reading
☐ Discourse processing
☐ Speech perception
☐ Prosody
☐ Other:
☒ Language production
☐ Cross-linguistic studies
☐ Second language processing
☐ Bilingualism
☐ Language acquisition
☒ Corpus-based studies
☐ Connectionist models
☐ Computational models and computer simulations
☐ Language in context

3. Name(s) of the author(s):
Present in the form Last name, First name, Middle initials.

1st: Frank, Austin F.
2nd: Jaeger, T. Florian
3rd:
4th:
5th:

4. Affiliation(s):
Please give affiliations of all authors.

1st: BCS, University of Rochester
2nd: BCS, University of Rochester
3rd:
4th:
5th:

5. Title of the presentation:
A Theory Of Choice In Production: Uniform Information Density Vs. Availability-Based Production

6. Abstract (max. 2600 characters with spaces, including references):

Recent work proposes an information-theoretical account of sentence production, Uniform Information Density [UID; Jaeger06, LevyJaeger06]: speakers structure their utterances so as to maintain a constant rate of information transmission (where INFORMATION(word)= -log PROBABILTY(word | context)). Initial evidence that speakers prefer UID comes from probability-conditioned pronunciation weakening, syllable/word shortening, and optional *that*-insertion. Optional *that* in (1) is more likely with less predictable complement clauses/clause onsets (we). UID competes with Availability-based Sentence Production [ASP; BockWarren85,FerreiraDell00], which claims that speakers insert that when the complement clause onset is not yet available for pronunciation. We present evidence favoring UID over ASP from two large-scale studies of morphosyntactic contractions in spontaneous speech (2-3).

(1) You think [(that) we should do this]?
(2) You're/are free to leave any time.
(3) He couldn't/could not do it.

For Study 1, we extracted 10,482 reduced and 6,042 full BEs (2) from a corpus of spontaneous speech. UID predicts that BE is more likely to be contracted where it is more predictable (and hence carries less information). In contrast, ASP predicts that BE is reduced

when the word following BE is predictable (and hence accessible). BE predictability and the predictability of the following word were estimated as conditioned on BE's potential host. E.g. for (2): P(ARE|you)=P('re|you) + P(are|you) and P(free|host=you).

We tested both hypotheses in the same multiple logistic regression model (bootstrapped to model speaker effects) with various additional controls (e.g. speechrate, phonological context, frequency of surrounding words, etc.). BE predictability affects speakers' choice between full and contracted BE more than any other factor (factor-removal: $\chi^2(1)=638.3$, $p<0.0001$): predictable BE is reduced more often. The predictability of the following word, however, has only a minimal effect ($\chi^2(1)=3.87$, $p<0.05$).
Study 2 extends these findings to 6,638 cases of reducible NOT (3). The redundancy of NOT is the most influential factor in speakers' choice. The predictability of the following word has no effect.

In conclusion, UID also affects morphosyntactic choice. UID captures speakers' preference to reduce forms that convey redundant information. Optional reduction is pervasive across as well as within languages; UID captures this important aspect of language production, while ASP misses it. We illustrate UID's predictions for several other typologically varied phenomena.