

**Instructions:**

Fill in this form.

Save it with the 1<sup>st</sup> 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](mailto:amlap2007@utu.fi).

The subject line of your email should be "submission".

# 13<sup>th</sup> 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"):

- |  |  |
|--|--|
| <input type="checkbox"/> Word recognition                      | <input checked="" type="checkbox"/> Language production                |
| <input type="checkbox"/> Lexical processing and representation | <input type="checkbox"/> Cross-linguistic studies                      |
| <input type="checkbox"/> Morphology                            | <input type="checkbox"/> Second language processing                    |
| <input type="checkbox"/> Parsing and interpretation            | <input type="checkbox"/> Bilingualism                                  |
| <input checked="" type="checkbox"/> Sentence comprehension     | <input type="checkbox"/> Language acquisition                          |
| <input type="checkbox"/> Reading                               | <input checked="" type="checkbox"/> Corpus-based studies               |
| <input type="checkbox"/> Discourse processing                  | <input type="checkbox"/> Connectionist models                          |
| <input type="checkbox"/> Speech perception                     | <input type="checkbox"/> Computational models and computer simulations |
| <input type="checkbox"/> Prosody                               | <input type="checkbox"/> Language in context                           |
| <input type="checkbox"/> Other:                                |  |

**3. Name(s) of the author(s):**

Present in the form Last name, First name, Middle initials.

1st: Jaeger, T. Florian

2nd:

3rd:

4th:

5th:

**4. Affiliation(s):**

Please give affiliations of all authors.

1st: BCS, University of Rochester

2nd:

3rd:

4th:

5th:

**5. Title of the presentation:**

**Rational Speakers: Speakers Help Processing Only When It Is Most Necessary**

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

The idea that human cognition is rational (Anderson90) has been highly influential in cognitive science. Recently this idea has also been applied to language production: Speakers are rational. According to Uniform Information Density (Jaeger06, LevyJaeger06; also Aylett99), speakers structure their utterances so as to avoid peaks and troughs in the information transmitted per time. Support for UID comes from reduction of redundant (i.e. predictable) optional material in phonetic and syntactic production: e.g. in (1), speakers are more likely to insert "that" in less predictable relative clauses (Jaeger06). UID has been shown to optimize comprehension (minimizing difficulty and/or maximizing communicative success) while keeping production efficient (LevyJaeger06). If speakers' decisions as to where to insert "that" indeed correlate with where processing would otherwise be extremely difficult, this would support the hypothesis of rational speakers. We test this hypothesis using a new approach that compares the distribution of "that" in production against results from comprehension.

We conducted a self-paced reading time study using 24 non-subject-extracted RCs (1) (and twice as many fillers) that were pseudo-randomly drawn from the parsed Wall Street Journal: 12 of them with that (1a) and 12 without (1b).

(1a) The way [RC that we've been managing... [Original: w/ that]

(1b) The ball [RC he hit wasn't a strike. [Original: w/o that]

The stimuli were systematically sampled so that there were representative of the corpus. For each stimulus, a matched stimulus was created with "that" if the original did not have "that" (1d) and vice versa (1c).

(1c) The way [RC we've been managing ... [Original: w/ that; Stimulus: w/o that]

(1d) The ball [RC that he hit wasn't a strike. [Original: w/o that; Stimulus: w/ that]

Overall "that" improves comprehension in the subject region (e.g. "he" in (1b,d); "we" in (1a,c);  $F_1(1,36)=25.6$ ,  $p<0.0001$ ;  $F_2(1,11)= 6.1$ ,  $p<0.04$ ; see also RaceMacDonald03). Interestingly, original and manipulated stimuli are overall read equally fast. That is due to an interaction with "that"-presence ( $F_1(1,36)= 17.9$ ,  $p< 0.0001$ ,  $F_2(1,11)= 7.0$ ,  $p< 0.023$ ): inserting "that" always significantly improves comprehension, but producers do NOT insert "that" all the time. Reading times on the original stimuli with and without "that" were statistically indistinguishable.

That means speakers only insert "that" when it is most necessary (when the RC onset is difficult). This is the first evidence that speakers' choices seem to be rational, as predicted by UID.