From Message to Syntax: Incremental Syntactic Planning *beyond* the Clause Level

Carlos Gómez Gallo (University of Rochester), T. Florian Jaeger (University of Rochester), and Ronald Smyth (University of Toronto)
cgomez@ling.rochester.edu

Most research on production focuses on planning below the clause level (segmental, lexical, and syntactic planning), whereas little is known about how speakers determine how to distribute information across multiple utterances. We present the first evidence that speakers' planning beyond the clause level is not limited to such cases of availability-based production [1,3,6] or macroproposition models [2]. **Instead, speakers have early access to at least some measure of clause complexity and they use that knowledge to limit the amount of information conveyed per clause by planning how many clauses will be used to realize the message.** This result suggests an extension of Uniform Information Density [5] beyond the clause level.

Consider a speaker planning to convey the message MOVE *theme* TO *location*. The speaker may either split the message into an explicit SELECT *theme* followed by MOVE *it* (2), or simply MOVE *theme* with an implicit SELECT (1). Because an explicit SELECT leaves additional time to plan the location expression, availability-based models would claim that the choice is driven by location complexity, but not theme complexity. If, on the other hand, speakers limit the information conveyed per clause, both higher theme and location complexity should correlate with a higher likelihood of splitting the message into two clauses.

To test this hypothesis, we analyzed all utterances from the Continuous Understanding Corpus [4], in which speakers direct listeners to move, rotate or paint objects on the screen. Objects and locations differed in complexity (shape, size, decoration, etc.). We used multiple logistic regression to assess how theme and location complexity (measured in log number of characters, underlined in (1-2)) influence clause production. Speakers were more likely to split up their message (2) when the theme and location were more complex (e^{β} =4.9, p<0.001; e^{β} =2.3, p<0.02, respectively). We also coded for givenness of the theme. Given themes carry less information than new ones, and are also more accessible. Based on the lower resource requirements for given themes, we predicted that they would have fewer clause splits, and this was confirmed (e^{β} =-2.12, p<0.04). Speakers also prefer two clauses when elaborating on a previously unclear MOVE (e^{β} =2.5, p<0.01, e.g. "a little more", "a little more"). Crucially, the effect of theme complexity remains highly significant.

We conclude that speakers decide early during production whether or not to split messages, and prefer to distribute high-information arguments across several clauses. This finding goes beyond availability-based production, [1,3,6], and also macro-proposition accounts, [2], and argues for a limited resource associated with clause planning.

From Message to Syntax: Incremental Syntactic Planning *beyond* the Clause Level

(1) **Conveying the entire message in one clause**
    Put [the square] [up into the Forest Hills]. (MOVE with implicit SELECT)

(2) **Splitting the message into two clauses**
    Take [the square with the heart on the corner] (explicit SELECT) and move [it] [up into Forest Hills]. (MOVE)

[1] Bock, J.K.; & Warren, R.K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition.* pp 47-67.

[2] Dell, G.; & Brown, P. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In D. J. Napoli & J. A. Kegl (Eds). *Bridges between psychology and linguistics*. pp. 111-129. Hillsdale, NJ: Erlbaum.

[3] Ferreira, V.S.; & Dell, G.S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40:296-340.

[4] Gomez Gallo, C.; Aist, G.; Allen, J.; Beaumont, W. de; Coria, S.; Gegg-Harrison, W.; Pardal, J.; & Swift, M. (2007). Annotating continuous understanding in a multimodal dialogue corpus. *DECALOG-07*, pp. 75-82.

[5] Jaeger, T.F. (2006). Redundancy and syntactic reduction in spontaneous speech. Ph.D. thesis, Stanford University, Stanford, CA.

[6] Levelt, W.J.M.; & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. Levelt (Eds.), *Crossing the boundaries in linguistics*. pp. 221-252. Dordrecht: Reidel.