Investigating the role of distribution in syntactic structures

Valerie J. Langlois (University of North Carolina at Chapel Hill)
valeriel@live.unc.edu

Comprehenders encounter a variety of syntactic structures in everyday life, whether through reading or spoken conversation. Some theoretical models of syntactic processing claim that comprehenders can acquire the frequency statistics of syntactic structures from exposure, which in turn leads to syntactic expectations (Levy, 2008; MacDonald et al., 1994; MacDonald & Thornton, 2009). These models imply that not only do comprehenders have implicit statistical knowledge of the relative frequencies of syntactic structures given a verb, but also that they can adapt to distributional changes. Infrequent structures (e.g. reduced relative clauses, such as *The soldiers warned about the dangers conducted the raid*) impose more difficulty as measured by reading time (MacDonald et al., 1994), and previous work has shown that this difficulty decreases with repeated exposure (Fine et al., 2013; Wells et al., 2009). In contrast, the role of distribution has only been investigated by correlating data from corpora and reading times (Gennari & MacDonald, 2009). Therefore, the current study asks the question: Do comprehenders keep track of the distribution of syntactic structures?

The current study investigates whether comprehenders acquire syntactic distributional information by directly manipulating the relative frequency of two syntactic structures: the *needs* structure (*The meal needs cooked,* a variation of *The meals needs to be cooked*) and the modifier structure (*The meal needs cooked vegetables*). As part of an American regional dialect, most people are relatively unfamiliar with the *needs* structure. Despite this unfamiliarity, comprehenders can rapidly adapt to the *needs* structure with enough exposure (Fraundorf & Jaeger, 2016; Kaschak & Glenberg, 2004). Critically, both structures are syntactically ambiguous until two words after *needs.* If comprehenders implicitly keep track of the distribution of structures that co-occur with the verb *needs*, then a distribution with a higher proportion of *needs* structures should result in less processing difficulty during disambiguation.

Methods: Participants were assigned to one of two conditions: the 80-20 condition (n=19) and the 40-60 condition (n=20). These numbers represent the relative percentages of the two syntactic structures (*needs* and modifier structure respectively). In the 80-20 condition, participants completed a self-paced reading task in which they read 20 *needs* structures, 5 modifier structures, and 55 unrelated filler sentences. Out of all the sentences that used the verb *needs*, 80% were the *needs* structures, and 20% were the modifier structures. Likewise, in the 40-60 condition, participants read 20 *needs* structures (40%), 30 modifier structures (60%), and 30 unrelated filler sentences. Modifier structures were presented at specific timepoints in the experiment, so that at any given *needs* structure, the distribution of *needs* to modifier sentences would be as close to the target distribution as possible. After reaching the end of the sentence, participants answered one comprehension question to ensure they read the sentence. Notably, participants in both conditions read precisely the same number of *needs* structures in the same order. Thus, if mere exposure drives facilitation, no difference is expected across conditions. In contrast, if comprehenders track the distribution of these two structures, then there should be a difference even when controlling for overall exposure.

Results: Following Fraundorf & Jaeger (2016) and the self-paced reading literature, reading times were corrected for word length, baseline reading speed, and task adaptation. The novel *needs* structures were analyzed at the disambiguating region. Reading times for the disambiguating region decreased faster for the 80-20 condition than the 40-60 condition (Fig.2), as illustrated by a significant interaction between condition and order (p = .01). There was a marginal main effect of condition (p = .1), and no main effect of order.

Conclusion: A higher proportion of *needs* sentences led to a faster rate of syntactic adaptation, independent of overall exposure. The difference in reading rate across the two distributions suggests that comprehenders are sensitive to the change in distribution. This shows that comprehenders can acquire syntactic distributional information, consistent with experience-based models of syntactic processing (e.g. MacDonald et al., 1994).

Table 1: Example sentence for each structure.

***Needs* structure:**	The fire needs stoked to keep it from burning out.
Modifier structure:	The meal needs cooked vegetables so the guests will be happy.

Table 2: Summary of model results at the disambiguating region for *needs* structures.

Model Parameters	Estimate	Std. Error	df	t-value	*p*-value
Intercept	20.402	6.711	33.300	3.040	0.00458
Condition	-14.502	8.482	35.363	-1.710	0.09608
logOrder	-9.262	8.021	40.099	-1.155	0.25504
Condition*logOrder	-23.923	9.555	687.102	-2.504	0.01252

Order of presentation was regressed out in the length-corrected reading time model. The DV (residual reading time) was kept in its raw form as log-transforming led to similar results.

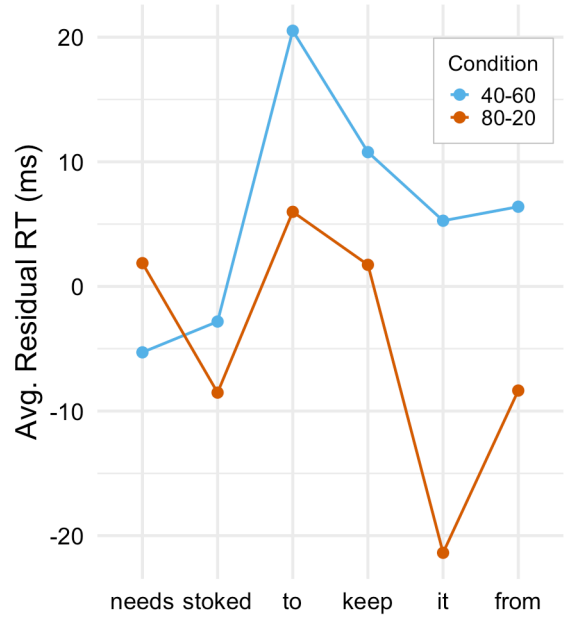

Figure 1: Average residual reading time for all novel *needs* structures, broken down by condition. Disambiguating region is two words after "needs" (e.g. "to").

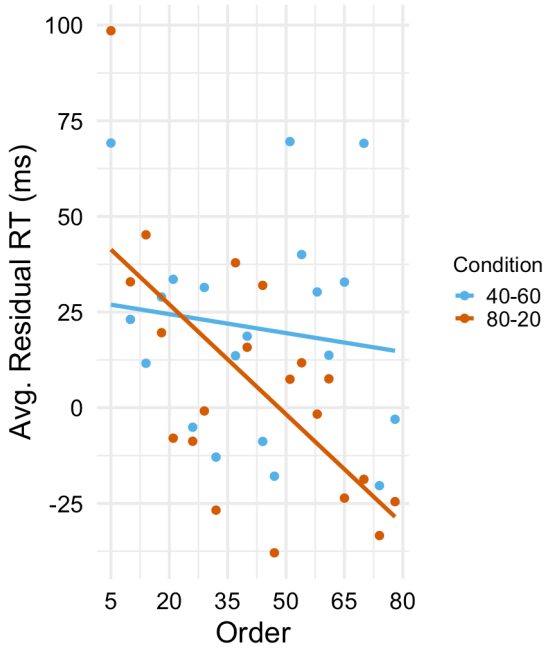

Figure 2: Average residual RT during the disambiguating region over the course of the experiment, broken down by condition.

References: Fine et al. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10). Fraundorf & Jaeger (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, *91*, 28–58. Gennari & MacDonald (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23. Kaschak & Glenberg (2004). This construction needs learned. *Journal of Experimental Psychology*, *133*(3), 450–467.Levy (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. MacDonald et al. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. MacDonald & Thornton (2009). When language comprehension reflects production constraints: Resolving ambiguities with the help of past experience. *Memory and Cognition*, *37*(8), 1177–1186. Wells et al. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58(2), 250–271.