Distribution matters: change in relative frequency affects syntactic processing

Valerie J. Langlois & Jennifer E. Arnold (University of North Carolina – 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). Yet theoretically, processing is specifically impacted by distribution – i.e., the relative frequency of target and competing structures. But the role of distribution has only been investigated by correlating data from corpora and reading times (Gennari & MacDonald, 2009). In the current study we provide the first experimental test of whether comprehenders keep track of the distribution of syntactic structures.

We investigate whether comprehenders acquire syntactic distributional information by directly manipulating the relative frequency of two competing syntactic structures: the dialectal *needs* structure and the modifier structure (Table 1). The dialectal *needs* structure is unfamiliar to most people (apart from those in Western Pennsylvania; Murray et al. 1996). Despite this unfamiliarity, comprehenders can rapidly adapt to the dialectal 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 *needs*, then a distribution with a higher proportion of dialectal *needs* structures should result in less processing difficulty during disambiguation, independent of overall exposure.

Methods: We used a 2x2 between-subjects design with two distributions and an ambiguous and unambiguous condition. 233 participants were assigned to one of two distributions (80-20 vs. 40-60) with either the dialectal structure or the standard structure. The numbers in each distribution represent the relative percentages of the two syntactic structures (dialectal/standard and modifier structure respectively). In the 80-20 distribution, participants completed a self-paced reading task in which they read 20 target *needs* (80%), 5 modifier structures (20%), and 55 unrelated fillers. Likewise, in the 40-60 distribution, participants read 20 target *needs* (40%), 30 modifier structures (60%), and 30 unrelated fillers. Modifier structures were presented at specific timepoints in the experiment, so that at any given target structure, the distribution of target to modifier sentences would be as close to the target distribution as possible. At the end of the sentence, participants answered one comprehension question to ensure they read the sentence. Notably, participants in both distributions read precisely the same number of target structures in the same order. Thus, if mere exposure drives facilitation, no difference is expected across the ambiguous conditions. In contrast, if comprehenders track the distribution of the dialectal and modifier structures, then there should be a difference even when controlling for overall exposure.

Results: Reading times were corrected for word length, baseline reading speed, and task adaptation. The target *needs* structures were analyzed at the same word (e.g. *before*). There was a significant three-way interaction between distribution, ambiguity, and order ($p<.05$, Fig.1); Reading times for the disambiguating word decreased faster for the ambiguous 80-20 condition than the 40-60 condition ($p<.05$), but not for the unambiguous conditions ($p=.93$).

Conclusion: A higher proportion of dialectal *needs* sentences led to a faster rate of syntactic adaptation, independent of overall exposure. The difference in reading rate across the two distributions in the ambiguous condition 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.

Dialectal structure:	The fire needs stoked to keep it from burning out.
Standard structure:	The fire needs *to be* 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 critical word (e.g. *to*) for target structures.

Model Parameters	Estimate	Std. Error	df	t-value	*p*-value
Intercept	10.546	4.402	27.329	2.396	0.02369
Distribution (80-20 vs. 40-60)	5.515	5.227	220.71	1.055	0.29256
Order	-17.117	5.547	18.022	-3.086	0.00637
Ambiguity (1 vs. 0)	31.078	5.227	220.7	5.945	< .001
Distribution*Order	-7.027	4.914	4155.581	-1.43	0.15285
Distribution*Ambiguity	-4.379	10.455	220.714	-0.419	0.67571
Ambiguity*Order	-10.738	4.914	4155.633	-2.185	0.02894
Distribution*Order*Ambiguity	-22.237	9.829	4155.704	-2.262	0.02373

Order of presentation (log-transformed) was regressed out in the length-corrected reading time model and centered in the final model. Distribution and ambiguity were contrast-coded, with 40-60 and ambiguity=0 as the reference level.

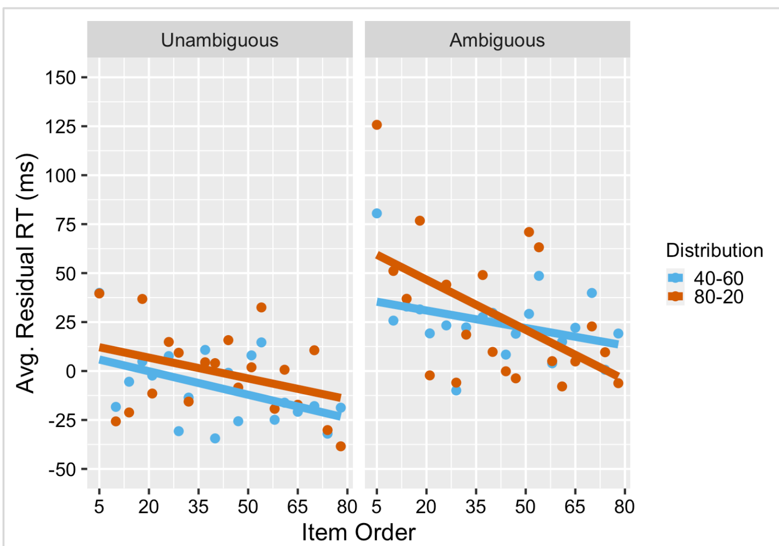

Figure 1: Average residual RT during the disambiguating region over the course of the experiment, broken down by distribution and ambiguity condition.

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.