Online Effort from a Minimalist Grammar Parser Improves RT Fit in RC Processing

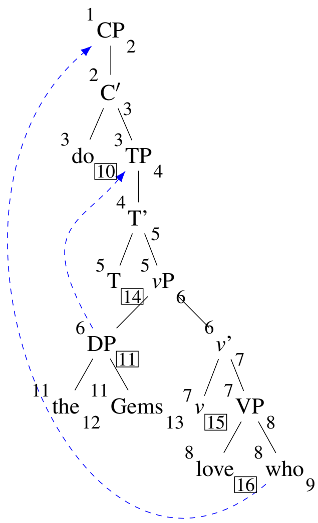
Aniello De Santo (University of Utah)

Overview. A parser for Minimalist grammars (MGs; Stabler 2013) has been shown to successfully model sentence processing preferences across an array of languages and phenomena when combined with complexity metrics that relate parsing behavior to memory usage (Gerth 2015; Graf et al. 2017; De Santo 2020:a.o.). This approach (henceforth: **MG Model**) provides a quantifiable theory of the effects of fine-grained grammatical structure on cognitive cost, and can help strengthen the link between generative syntactic theory and sentence processing. However, work on it has focused on modeling *off-line* asymmetries. Here, we extend this approach by showing how measures of effort that explicitly consider minimalist-like structure-building operations improve our ability to account for word-by-word (*online*) behavioral data.

The MG Model. The model we adopt links structural details to processing load by associating the stack states of a (deterministic) top-down parser (Stabler 2013) to memory burden (Kobele et al. 2013). This parser is *string-driven*: when encountering a displaced word (e.g., "*who*"), it prioritizes finding a path to its base position. In this abstract, memory usage is then measured based on how long a node is kept in memory through a derivation, tracking how the derivational operations interact with fine-grained structural details to affect linear word order (*Tenure*). The annotation schema of Fig. 1 captures how the parser's tree traversal strategy affects memory: the superscript (index) of a node n encodes the moment n was predicted and put in memory. The subscript (outdex) encodes the moment n is confirmed and frees up memory. Tenure for n is $outdex(n) - index(n)$: e.g. *Tenure*$(do) = 10 - 3 = 7$. While past work has leveraged offline metrics estimating effort for a full derivation, we can derive online measures by extracting Tenure values for every (pronounced) lexical item (Fig. 1).

Evaluating Tenure Online. Offline subject/object relative clause (SRC/ORC) asymmetries have been extensively probed with the MG Model (Graf et al. 2017; De Santo 2020). Because of this, we ask whether structure-building effort as captured by Tenure improves model fit to the self-paced reading data made available for English SRCs/ORCs in the Syntactic Ambiguity Processing Benchmark (Huang et al. 2024), beyond established expectation-based predictors. First, we fit a baseline linear mixed-effects model to the RTs, with several lexical control predictors. We then add to the baseline model surprisal predictors, fitting two models with surprisal values derived either from an LSTM (Gulordava 2018) or GPT-2 small (Radford et al. 2019). Then, we compute via the MG model word-by-word Tenure values for derivations built for RC items in the benchmark. The MG trees follow standard generative assumptions for the main clause of each sentence, and a wh-movement analysis for the structure of RCs (Chomsky 1977). Finally, we fit two models adding these MG Tenure values to the two surprisal models. The best performing model was the *GPT-surprisal* + *Tenure* model (Table 1), showing that taking Tenure into account significantly improves model fit to RT data. In particular, we found that Tenure of both the current word and the preceding two words is associated with significantly slower RTs independently of surprisal (Table 2).

Results and Discussion. Our results show that predictors relying on explicit structure-building operations improve our ability to model word-by-word RTs, beyond the contribution of surprisal measures — adding support to the cognitive relevance of transparent structure-building measures, and to the use of the MG Model in investigating the interaction of generative syntax and human sentence processing. Additionally, the model's sensitivity to fine-grained grammatical assumptions implies that analytical choices have a significant impact on the derived Tenure values. Thus, future work could exploit online behavioral data to distinguish competing syntactic proposals (and, potentially, different syntactic formalisms) based on their psycholinguistic predictions, thus clarifying how/which aspects of sentence structure modulate processing difficulty.



	Who	do	the	Gems	love
Tenure	1	7	1	2	8

Figure 1: Example of an MG derivation tree for *Who do the Gems love?* with annotated parse steps, and tenure values for pronounced lexical items. Unary branches indicate movement landing sites.

	df	AIC	BIC
Baseline	14	1047981	1048110
LSTM Surprisal	19	1047506	1047681
GPT Surprisal	19	1047414	1047589
LSTM Surprisal + Tenure	23	1045549	1045761
GPT Surprisal + Tenure	**24**	**1045493**	**1045714**

Table 1: Model Comparison.

	RT			
Predictors	*Estimate*	*Std. Error*	*t value*	
(Intercept)	410.8253928	5.3006171	77.5052004	***
Tenure	5.5809141	1.2603398	4.4281028	***
Tenure $i-1$	12.1147670	1.4416750	8.4032578	***
Tenure $i-2$	5.0687843	0.9777471	5.1841468	***
Surprisal	13.4585782	1.9321063	6.9657547	***
Surprisal $i-1$	11.6738518	1.7103038	6.8256014	***
Surprisal $i-2$	1.5617456	1.8103692	0.8626669	
Word Position	4.7241724	1.1106569	4.2534938	***
logfreq	0.9923497	1.9407082	0.5113338	
length	18.1574795	2.0805941	8.7270649	***
logfreq $i-1$	-0.4084933	1.9069223	-0.2142160	
length $i-1$	8.5549350	2.1022739	4.0693723	***
logfreq $i-2$	-2.9067570	2.0909322	-1.3901728	
length $i-2$	3.4984582	2.0355144	1.7187097	
logfreq:length	0.8208	1.4693	0.559	
logfreq $i-1$:length $i-1$	-4.4242	1.6809	-2.632	**
logfreq $i-2$:length $i-2$	-0.5443	1.6193	-0.336	

Table 2: Lmer Summary for best fitting model (GTP Surprisal + Tenure).

Chomsky. 1977. On wh-movement. **De Santo.** 2020. Structure and memory: A computational model of storage, gradience, and priming. **Gerth**. 2015. Memory limitations in sentence comprehension: A structural-based complexity metric of processing difficulty. **Graf et al**. 2017. Relative clauses as a benchmark for Minimalist parsing. JLM. **Gulordava et al**. 2018. Colorless green recurrent networks dream hierarchically. ACL. **Huang et al**. 2024. Large-scale benchmark yields no evidence that language model surprisal explains syntactic disambiguation difficulty. JML. **Kobele et al**. 2013. Memory resource allocation in top-down minimalist parsing. FG. **Radford et al**. 2019. Language models are unsupervised multitask learners. **Stabler**. 2013. Two models of minimalist, incremental syntactic analysis. TiCS.