Surprisal and Event Structure in Reading Garden Path Sentences

Ryan King (NYU) and David Townsend (Montclair State University)
ryan.king@nyu.edu

We ask whether probabilistic factors (as measured by surprisal) fully account for sentence processing difficulty, or whether event structure factors (here, telicity) contribute to anticipation-based processing effects. The event structure hypothesis [9] proposes that telic verbs elicit anticipation of an affected patient that may appear in a direct object (*The dog halted the burglar*) or in the subject (*The dog halted*), facilitating assignment of the subject to the patient role. The difficulty of the reduced version of *The dog (that was) {raced / halted} by the owner jumped the fence* when the initial verb is atelic (*raced*) supports the event structure hypothesis [9, 10].

We calculated surprisal from the transitional probability of a *by*-phrase in the context of an atelic vs. telic verb [4, 6]. If surprisal is the sole bottleneck in processing, difficulty will increase as surprisal increases [5, 11]; event structure will explain no additional variability in difficulty.

We recorded eye fixation times as 417 participants read garden path sentences adapted from Malaia et al. [7, 8]. We used 100 different initial verbs in 364 sentences with reduced (RC) or unreduced (UC) clauses for a total of 7,241 trials. Half of the initial verbs (*v1*) were atelic and half were telic. Both noun phrases were animate. Using the 14 billion-word iWeb corpus [3], we calculated Surprisal (S) as $-\log_2(F(v1 + by + the + \text{NOUN}) / F v1)$, where F refers to frequency, italics refer to lexical items (e.g., *raced by the*) and capitals refer to part of speech (*owner* is a noun). S was unrelated to log frequency of *v1* ($r = -0.008$, $p > 0.10$) and of the noun in the *by*-phrase (*n2*, $r = 0.0001$, $p > 0.10$). Atelic and telic verbs did not differ in S, F, or length, all *p*s > 0.10.

We modeled first pass time (FPT), regression path duration (RPD) and total time (TT) [12] for the *by*-phrase (*by the owner*) and the main verb (*v2*, *jumped*). FPT tends to assess lexical retrieval while RPD and TT tend to assess text integration [2]. We deleted reading times 4 SD greater than the mean (< 1% of data in each region) and centered all independent variables. Fixed effects included log frequency of *v1, n2* and *v2,* S, Clause Type (CT), Telicity (Tel), and the CT*S, CT*Tel, and CT*Tel*S interactions. Modeling proceeded from the most to the least complex fixed effect structure, dropping non-significant interactions from further models. Due to lack of convergence, the models used random intercepts for participants and items. The model results appear in Table 1 and the reading time data in Table 2.

The *by*-phrase showed a CT*Tel*S interaction in FPT primarily due to FPT decreasing with S in atelic UCs (atelic RC: $\beta = -10.767$, se = 4.933, $p = 0.0297$; atelic UC: $\beta = -25.663$, se = 7.007, $p < 0.001$; telic RC: $\beta = 9.404$, se = 3.142, $p = 0.003$; telic UC: $\beta = 4.432$, se = 5.574, $p = 0.428$). The *by*-phrase also showed a CT*S interaction in TT with TT increasing with S in RCs ($\beta = 29.854$, se = 4.783, $p < 0.001$) but not in UCs ($\beta = 13.075$, se = 6.285, $p = 0.378$). The main verb showed a CT*S interaction in RPD with duration increasing with S in RCs ($\beta = 14.218$, se = 4.106, $p = 0.001$) but not in UCs ($\beta = 3.509$, se = 5.720, $p = 0.539$). The main verb also showed a CT*Tel*S interaction in TT in which surprisal effects appeared in all conditions except for telic verbs in UCs (Figure 1; atelic RC: $\beta = 21.110$, se = 6.672, $p = 0.002$; atelic UC: $\beta = 21.919$, se = 8.792, $p = 0.014$; telic RC: $\beta = 13.219$, se = 4.598, $p = 0.004$; telic UC: $\beta = -6.120$, se = 7.173, $p = 0.395$).

Detailed analysis of eye movement patterns can clarify the interactions between telicity and surprisal. In general, these interactions indicate that probabilistic factors do not fully account for processing difficulty. In particular, the clause type, telicity and surprisal interaction in total time on the main verb indicates that the processing system has not fully updated the interpretation on the *by*-phrase and is still working out an interpretation when reading the main verb. An exception is a sentence with a telic verb in an unreduced clause, which shows no surprisal effect on the main verb. In general terms, the results suggest a model with two systems [e.g., 1] in which co-occurrence probabilities elicit anticipations that subsequently are integrated with event structure. Research in progress examines whether surprisal based on a language model of higher quality explains the pattern of eye movements while reading garden path sentences.

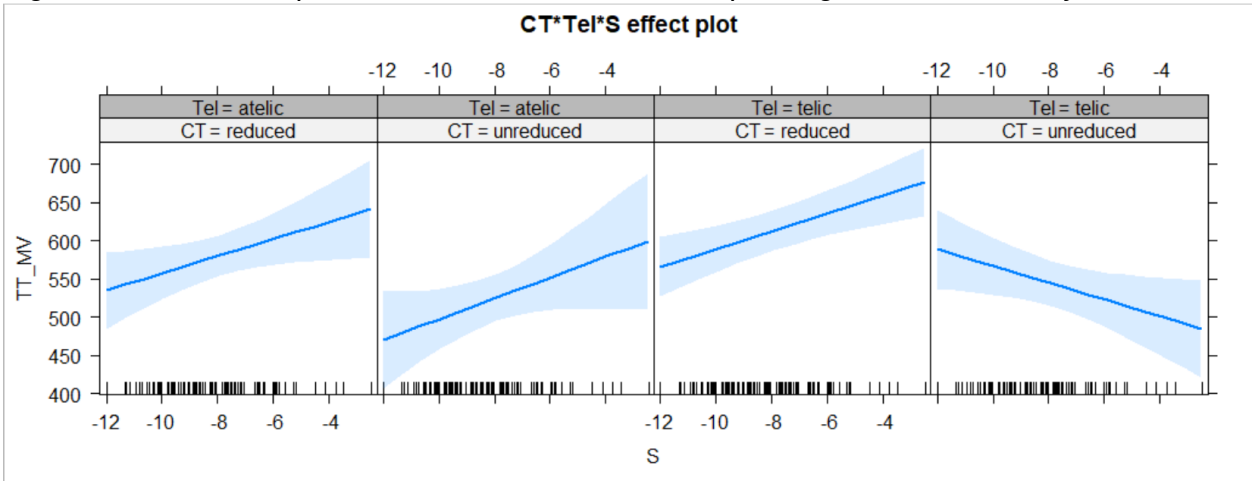
Table 1 Likelihood Ratio Tests (bold font indicates p < 0.05)

	FPT		RPD		TT	
By-Phrase	χ^2	*p*	χ^2	*p*	χ^2	*p*
CT	14.249	**0.0002**	121.01	**<0.0001**	390.64	**<0.0001**
S	0.4289	0.5125	10.473	**0.0012**	26.206	**<0.0001**
CT*S	6.2118	**0.0448**	3.3887	0.1837	11.75	**0.0028**
CT*Tel	0.1545	0.6943	2.6247	0.1052	0.5107	0.4748
CT*Tel*S	27.712	**< 0.001**	5.4854	0.0644	5.5154	0.0634
Main Verb						
CT	5.4299	**0.0198**	16.442	**<0.0001**	45.755	**<0.0001**
S	0.0002	0.9897	7.1697	**0.0074**	9.7153	**0.0018**
CT*S	0.4855	0.7845	6.5662	**0.0375**	8.690	**0.0032**
CT*Tel	0.3512	0.5534	0.7512	0.3861	0.3808	0.5372
CT*Tel*S	2.1503	0.3413	2.2383	0.3266	7.748	**0.0208**

Table 2 Means (ms) by Reading Time Measure, Region, CT, and Tel

	FPT		RPD		TT	
By-Phrase	Reduced	Unreduced	Reduced	Unreduced	Reduced	Unreduced
Atelic	486	522	753	609	1171	890
Telic	520	548	714	601	1191	907
Main Verb						
Atelic	312	323	506	447	583	542
Telic	315	323	505	465	614	560

Figure 1 Effects of Surprisal on TT for the Main Verb Depending on CT and Telicity



References

[1] Brothers & Kuperberg 2021. *JML* [2] Clifton et al. 2007. In *Eye Movements: A Window on Mind and Brain*. [3] Davies 2018. https://www.english-corpora.org/iWeb/. [4] Frisson et al. 2005. *JEP:LMC*. [5] Levy 2008. *Cognition*. [6] MacDonald & Shillcock 2003. *Psychological Science*. [7] Malaia et al. 2009. *Brain & Language.* [8] Malaia et al. 2012. *J Psycholinguist Res.* [9] O'Bryan 2003.U of Arizona dissertation. [10] O'Bryan et al. 2013. In *Syntax and Its Limits*. [11] Smith & Levy 2013. *Cognition*. [12] Staub https://blogs.umass.edu/eyelab/software/