Multitasking Events in English: Cross-task Conceptual Replications and Re-evaluations

Lisa Levinson, Rohan Raju, Lucy Yu-Chuan Chiang, & Sid Bhusan (University of Michigan)

What role do lexical representations of event structure play in the temporal dynamics of sentence processing beyond relative event-type probabilities, and to what extent can we detect these effects in behavioral reading and reaction times? Do we find complexity costs for analytically more complex representations, or only for special cases such as coercion?

Background: Previous psycholinguistic work has found processing correlates of lexical semantic complexity in verbs [1,2,3,4,5], as well as event-associated coercion effects [6,7] which have been argued to be independent of probability and prediction. Recent work however has focused primarily on broader effects of prediction and memory as the central mechanisms impacting processing time and difficulty. Since much of the work on lexical event representation and coercion was conducted prior to the availability of higher-powered online data collection and contemporary methods of statistical analysis and language-model generated lexical predictions, here we aim to replicate several studies of event processing in English and, where effects are replicated, model these along with GPT2-based surprisal values to evaluate the extent to which these effects appear to be independent of lexical prediction (along the lines of [7,8]).

Current Studies: Three studies were conducted using (sub)sets the same stimuli, adapted to three behavioral tasks. The materials for the Maze task [9] experiment included stimuli from three prior studies, each study serving as filler for the others along with other unrelated fillers (Table 1), counterbalanced as in the originals. One set was 44 pairs from Gennari and Poeppel 2003 (GP03) exp 2, testing for a cost of eventivity over stativity at the verb in self-paced reading (SPRT). The second were sentences from exps. 2 and 3 in McKoon and Love 2011 (ML11) (12 pairs), testing for slower reaction times for sentences with change-of-state over activity verbs, or to the verb itself in the stop-making-sense (SMS) task. The third set was 30 quadruplets from Brennan and Pylkkänen 2010 (BP10) [10] exp 1, where SPRT was used to test both for effects of lexical complexity (subject experiencer vs object experiencer psych verbs) and aspectual coercion (telic over atelic SubjExp). Only the GP03 and ML11 stimuli were used for the SPRT and SMS replications. All experiments were conducted online using the Ibex platform [11]. Lexical surprisals were generated for all items with GPT-2 medium [12].

Exp 1 SPRT: 75 American English speakers from Prolific completed a self-paced moving window reading task. **Exp 2 SMS:** 131 American English speakers completed the SMS task for course credit. **Exp 3 Maze:** 48 American English speakers participated via Prolific. Maze alternatives were generated using the A-maze tool [13].

Results and Conclusions: For all experiments, log RTs were modelled with fixed effects for the relevant conditions by stimuli set and random effects for participants and items. Trials with incorrect responses to the task or comprehension question were excluded. Neither GP03 nor ML11 replicated in any task at critical (see Table 2) or spillover regions. Both effects of BP10 replicated in the maze task, and model comparison suggests that both the linguistic conditions (LRT $p < .001$) and surprisal (LRT $p < .001$) significantly improve fit over baselines. Given the greater power of these studies compared to the originals, the lack of replication of several of these classic effects of event complexity calls for further examination, and suggests that lexical complexity associated with events may not have a reliable, measurable, and distinct impact in these behavioral measures. However the replication of BP10's lexical semantic and coercion effects even beyond measures of lexical surprisal suggests that some such effects may be of a larger and more detectable magnitude and distinct from effects of lexical predictability.

Table 1: Sample Stimuli

Study	Conditions	Examples
McKoon & Love 2011	Break verbs	The workmen **chipped** the tiles.
	Hit verbs	The workmen **banged** the tiles.
Gennari & Poeppel 2003	Stative verbs	The young detective **disliked** his senior partner.
	Eventive verbs	The young detective **inspected** the crime scene.
Brennan & Pylkkänen 2010	Simple	According to the review the critic **hated** the popular actress.
	Coercion	Within moments the critic **angered** the popular actress.
	LexSem	According to the review the critic **angered** the popular actress.
	(ObjExp, telic)	Within moments the critic **hated** the popular actress.

Table 2: Statistical Model Estimates

Stimuli Set	Task	Level	Est.	SE	p-value
ML11	SPRT	change-of-state	.017	.02	> .1
GP03	SPRT	eventive	.006	.013	> .1
ML11	SMS	change-of-state	.014	.03	> .1
GP03	SMS	eventive	.003	.02	> .1
ML11	Maze	change-of-state	.015	.04	> .1
GP03	Maze	eventive	.040	.026	> .1
BP10	Maze	lexsem (ObjExp)	.112	.03	< .001
BP10	Maze	coercion (telic)	.097	.02	< .001

Selected References: [1]McKoon, G., & MacFarland, T. (2000). *Language*, vol. 76. [2]McKoon, G., & Macfarland, T. (2002). *Cogn Psych*, vol. 44. [3]Gennari, S., & Poeppel, D. (2003). *Cognition*, vol. 89. [4] Levinson, L., & Brennan, J. (2016). Morphological Metatheory, Vol. 229. [5]McKoon, G., & Love, J. (2011). *Lang & Cogn*, vol. 3. [6]Brennan, J., & Pylkkänen, L. (2008). *Brain & Lang*, 106. [7]Delogu, F., Crocker, M. W., & Drenhaus, H. (2017). *Cognition*,161. [8]Levinson, L. (2023). *Exp in Ling Meaning*, 2. [9]K. I. Forster, C. Guerrera, and L. Elliot (2009). *Behav. Res. Methods*. [10]Brennan, J., & Pylkkänen, L. (2010). *LCP*. [11]Drummond, A. (2013). *Ibex farm*.[12]Radford, A., & others. (2019). OpenAI Blog, 1(8), 9. [13]V. Boyce, R. Futrell, and R. P. Levy (2020) *J. Mem. Lang.*, vol. 111.