

Relevance vs similarity: Bayesian Pragmatics provides the best quantitative model for EEG and Cloze data on contextual modulation in a predictive completion task

Markus Werning and Matthias Unterhuber

In language comprehension, a preceding discourse can affect a listener's implicit probabilistic predictions on the completion of a discourse (Nieuwland & Van Berkum, 2006). Presupposing that a concrete noun (n) often has an agentive lexical component (Pustejovsky, 1995), which refers to the typical way the denoted object (e.g., *cream*) is brought about (*whip*), we compare three models of how context (c) can modulate this lexical component and so change probabilistic predictions on a subsequent action verb (v), i.e., $P_n(v|c)$. In a predictive completion task (Werning & Cosentino, 2017) with the 2x2 design $\{+ALex, -ALex\} \times \{AStdCtx, ANewCtx\}$ (see Table 1), we contrast the continuation after the noun (e.g., *cream*) with verb expressing either the noun's agentive lexical component (*whip*, +ALex) or an alternative way of bringing about the object (*draw*, -ALex). We combine this contrast with either a standard context (AStdCtx) or a context that suggests the new way of bringing about the object (ANewCtx). To control for the effects of semantic similarity and relevance, we hold the semantic similarities between the verb and the context sentence constant over all four combinations of conditions, and vary the relevance of the context (AStdCtx, ANewCtx) for the action described by either verb (+ALex, -ALex) as well as the semantic similarity between the noun and the verb. We compare the explanatory power of the three models with regard to both Cloze values (Experiment 1) and ERP amplitudes on the verb (Experiment 2) in a multiple linear regression analysis.

The Semantic Similarity Model presupposes that the only predictor of a word succeeding a discourse is the semantic similarity between the word and the discourse. With the verb's v semantic similarity to the context sentence c held constant, the predictive probability $P_n(v|c)$ should be a strictly increasing function f_n^+ only of the semantic similarity between v and the immediately preceding noun n . **The Relevance Model** takes the only predictor to be that, when speakers choose their utterances, they aim to maximize relevance between context c and verb v . $P_n(v|c)$ is thus modelled as a strictly increasing function g_n^+ of the relevance for c to v . **The Bayesian Pragmatic Model** accounts for the rational cooperation between speaker and listener using Bayes's Theorem. The speaker's intention is to attribute, to the discourse's protagonist, an action (denoted by v) to be performed on a given object (denoted by n). The speaker embeds this action in a situation that s/he introduces by context c . The speaker has to choose c to let action $[[v]]$ appear relevant, such that the likelihood $P_n(c|v)$ is a function g_n^+ of the relevance between c and v . The prior probability of the speaker's intention to attribute $[[v]]$ to be performed on $[[n]]$, $P_n(v)$, is a function, f_n^+ , of the semantic similarity between v and n .

Model Predictions. Both, Cloze values and the amplitude of the N400 component, can be assumed to correlate with surprise, i.e., the negative logarithm of the predictive probability, $-\log(P_n(v|c))$. Linear approximations of the functions f_n^+ and g_n^+ lead to the predictions in Table 2.

Method. The material was in German and consisted of 40 quadruples of the 2x2 design described above (160 stimuli in total) and 40 distractors. Semantic similarity was determined in a distributional semantics framework by GloVe (Pennington et al., 2014).

Experiment 1. A forced choice Cloze test (N=35) was performed, where subjects could choose for each context (AStdCtx, ANewCtx) between +ALex, -ALex and four randomly chosen alternative

agentive verbs. Relevance ratings were elicited (N=38) by asking participants to rate the plausibility of the context given the verb on a 7-point Likert scale.

Experiment 2. The EEG study used the same material as Experiment 1 (N=26). The ERPs were measured on the critical verb (+ALex, -ALex; see Figure 2). The N400 component was determined in the interval 400–500ms for six central-posterior electrodes.

Results and Discussion. Table 2 compares the three models for both Cloze probabilities and pooled N400 amplitudes as predicted variables. In both datasets, the Bayesian Pragmatic Model (correlation with data shown in Figure 1) is the clear winner, having the lowest value in the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). Both criteria take into account the different numbers of predictors.

References

Pennington, J. et al. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (pp. 1532–1543). Stroudsburg, PA: Association for Computational Linguistic.

Werning, M., & Cosentino, E. (2017). The Interaction of Bayesian Pragmatics and Lexical Semantics in Linguistic Interpretation: Using Event-related Potentials to Investigate Hearers’ Probabilistic Predictions. In G. Gunzelmann et al. (Eds.), (pp. 3504–3509). Proceedings of the 39th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–1111. <https://doi.org/10.1162/jocn.2006.18.7.1098>

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

	+ALex	-ALex
AStdCtx	Maria prepares a cake buffet to impress her friends and makes ready everything necessary for it. She is already about cream to whip.	Maria prepares a cake buffet to impress her friends and makes ready everything necessary for it. She is already about cream to draw.
ANewCtx	Maria practices for a picture of a cake buffet and uses her notebook for her preliminary study. She is already about cream to whip.	Maria practices for a picture of a cake buffet and uses her notebook for her preliminary study. She is already about cream to draw.

Table 1. Example of experimental stimuli, translated from German to English. The word order of the translation follows the word order of the original.

Predicted Variable: Cloze Probabilities (Experiment 1)										
Model	Pred. Prob.: $P_n(v c)$	Surprise: $-\log(P_n(v c))$	N	df	RMSEA	r	r_{adj}	BIC	AIC	
Bayesian Pragmatic	$K \cdot g_n^+(Rel_n(c, v)) \cdot f_n^+(S(v, n))$	$a Rel_n(c, v) + b S(v, n) + k$	160	157	.22	.850*	.848*	-16.40	-25.64	
Relevance	$g_n^+(Rel_n(c, v))$	$a Rel_n(c, v) + k$	160	158	.23	.839*	.838*	-11.12	-17.27	
Semantic Similarity	$f_n^+(S(v, n))$	$b S(v, n) + k$	160	158	.40	.223*	.209*	175.54	169.38	
Predicted Variable: EEG Amplitude (400–500ms, Experiment 2)										
Bayesian Pragmatic			160	157	2.10	.451*	.439*	703.82	694.59	
Relevance			160	158	2.15	.401*	.394*	707.02	700.87	
Semantic Similarity			160	158	2.28	.246*	.234*	725.02	718.87	

Table 2. Model predictions and comparisons for the regression analysis of Cloze probabilities and pooled EEG amplitudes. *p < .001.

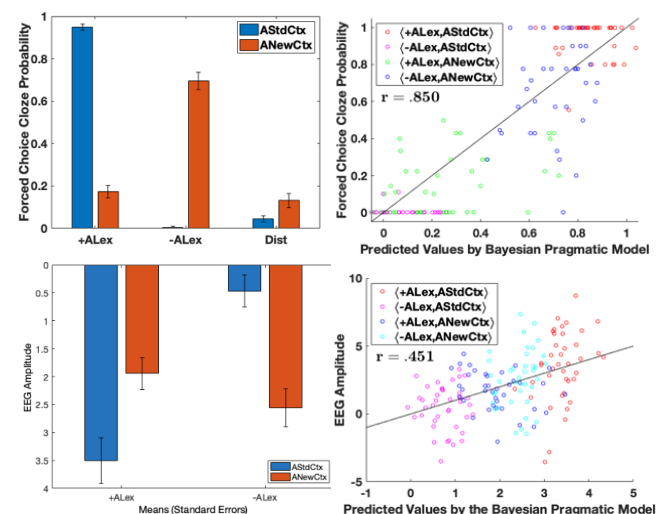


Figure 1. Top: Forced Choice Cloze Probabilities and Scatterplot for the Bayesian Pragmatic Model (Experiment 1). Bottom: EEG Amplitudes (400–500ms, μV) for the Pooled Central-Posterior Electrodes (CP1, CPz, CP2, P1, Pz, P2, Dist: Distractors) and Scatterplot (Experiment 2).

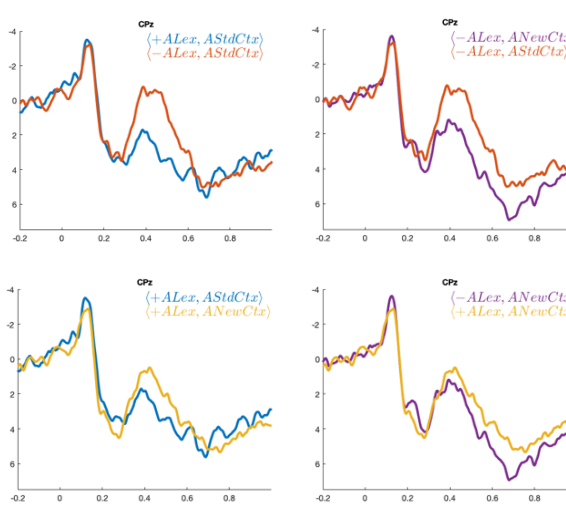


Figure 2. Base-line corrected ERPs measured on the CPz (μV).