Multi-word bottom-up effects in the visual world paradigm

Anuenue Kukona\(^1\) and Whitney Tabor\(^2\)

\(^1\)University of Dundee, \(^2\)University of Connecticut & Haskins Laboratories

contact: a.b.bakerkukona@dundee.ac.uk

Listeners show simultaneous anticipatory and lexically-based bottom-up effects in predictive contexts in the visual world. Kamide et al. (2003) found that listeners hearing sentences like “The girl will ride…,” with scenes that included a verb-predicted carousel and motorbike, anticipated the carousel predicted by the context (“girl” + “ride”). However, listeners also fixated the motorbike more with “(The girl will) ride…” vs. “taste…” Relatedly, Kukona et al. (2011) found that listeners hearing sentences like “Toby arrests the…,” with displays that included Toby, and a patient (crook) and agent (policeman) of the verb, anticipated the patient predicted by the context (active sentence + “arrest”). However, listeners also fixated the agent more than unrelated distractors (gardener). In both cases, listeners activated items (motorbike/policeman) predicted by the bottom-up verb constraints (“ride”/“arrest”), although these bottom-up constraints conflicted with the context (“girl”/active sentence). These results suggest that in addition to anticipating predictable items that fully satisfy the context, language users also simultaneously interpret the language signal based on context-independent, bottom-up (verb) input.

One interpretation of these results is that they support bottom-up, emergent theories of language processing that do not strictly enforce the coherence of language representations \([1,2,6,7]\). However, an alternative interpretation is that these results reflect independent lexical (bottom-up) vs. syntactic (anticipatory) processes \([3]\). To distinguish between these hypotheses, we tested for multi-word bottom-up effects in anticipatory contexts: such effects are unexpected given lexical-syntactic independence because syntactic/multi-word processes are hypothesized to have full access to contextual information.

**Methods.** Listeners (N = 33) heard 16 sentences like “The boy will eat the spotted, white cake” while viewing displays with items like: spotted, white cake; spotted, brown cake; striped, white cake; striped, brown cake; spotted, white car; spotted, brown car; striped, white car; and striped, brown car.

**Predictions.** We focus on verb-inconsistent items (cars, which are not predicted by “eat”), which allow us to distinguish the hypotheses. Bottom-up theories claim that the emergent, multi-word entity “spotted, white” is reinforced via feedback at the expense of entities like “spotted, brown,” “striped, white,” and “striped, brown,” predicting a preponderance of fixations to the spotted, white car relative to the other cars. Lexical-syntactic independence predicts no such interaction.

**Results.** Proportions of fixations to the verb-inconsistent items are plotted in Fig. 1. During “spotted,” there were reliably more fixations to the spotted, white/brown cars than the striped white/brown cars; during “white,” there were no reliable effects; and during “cake,” there was a reliable interaction, revealing more fixations to the spotted, white car than the remaining cars.

**Conclusions.** We found evidence for multi-word bottom-up effects in anticipatory contexts. These results are consistent with emergent theories, which assume that large-scale linguistic representations form via the ganging of local word structures. Non-feedback models, which form tree-structures on a strictly information-driven basis, have trouble simultaneously
predicting time-locked anticipatory effects [4,5], and the delay we observe in the appearance of the significant interaction. The emergent model predicts a delay because ganging depends on feedback, which in the case of input-inconsistent structure (cars), takes longer to develop than input-consistent anticipation effects.

Fig 1. Average proportions of fixations over time to verb-inconsistent items (e.g., cars; A), and average proportions of trials with fixations during “spotted,” “white,” and “cake” (windows begin at the offset of the relative word, and extend for 200 ms; B) for the example sentence “The boy will eat the spotted, white cake.”

References: