Towards a psychological evaluation metric for semantic representations

Frank Mollica (mollicaf@gmail.com, Rochester) & Edward A. Gibson (MIT)

When people read sentences, they update their mental model of the world via inferences. While there has been much progress identifying when and which inferences are made[1-3], one fundamental question in comprehension remains: What is the semantic/conceptual representation that is being updated? Formal semantics offers many representational schemes[e.g,4] based on simplicity and expressive capacity. Motivated by human reasoning, recent proposals from the concept literature suggest that concepts are best represented as probabilistic programs[5]. Implicitly, models of memory[6] are baseline domain-general encoding schemes for semantic content. One hurdle to comparing these varied schemes is identifying an impartial evaluation metric.

In this abstract, we propose a framework for evaluating semantic representation schemes independent of their original motivation. Fundamentally, representation schemes are encodings of semantic information. Under an information theoretic lens, these schemes share certain properties: 1) They are compressions (some lossy) of semantic information; 2) asymmetries in the priors of source and receiver sometimes result in only the approximate transfer of meaning; and 3) in channel limited systems (like humans), there is a utility function specifying which information to privilege when encoding. These three properties correspond to three observed behaviors of comprehenders: 1) retellings of complicated events are simplified[7]; 2) humans have false memories based on their priors[8-9]; and, 3) the information that humans forget when reading is non-random reflecting utility[10].

We provide exploratory evidence that both linguistic (i.e., argument/adjunct-hood) and conceptual properties of events (i.e., animacy and agentivity) influence recall and thus are factors likely included in comprehenders' utility functions. Specifically, we expect greater recall for arguments (constituents required by a verb) than adjuncts (constituents not required by a verb) because remembering an event serves as a cue for its arguments. Following ERP evidence highlighting the early online processing of animacy[11], we expect greater recall for animate arguments than inanimate arguments. Lastly, event recognition paradigms reveal privileged processing of agents[12]. Therefore, we expect greater recall for events with agents than for events without agents.

Following [13-14], participants ($N=160$) were instructed to read a one-paragraph story twice and recalled the story after a 10 min delay. They were instructed: "Please write down as much as you can about the passage that you read at the beginning of the experiment. You can use your own words but it is important that your version stay true to the original." Each participant read one story consisting of on average 14 events. The stories were excerpts from fiction ($N=10$), non-fiction books ($N=3$) and constructed paragraphs ($N=6$), which varied the agency of events (e.g., *the wind/the janitor closed the door...*). We used [15] to extract events from the stories and annotate argument/adjunct-hood according to [16], hand correcting errors. Animacy and agency were annotated by hand. Responses were hand coded as either: Correct or Error/Omitted.

Consistent with [14], participants recall on average 33% of propositions. Figure 1 presents the results for one example story. To evaluate our predictions, we conducted mixed effect logistic regression models. As can be seen in Figure 2, our predictions were confirmed. Participants recall more arguments than adjuncts (Fig 2a; $z=7.4,\ p<0.05$). Participants recall more animate arguments than inanimate arguments (Fig 2b; $z=4.5,\ p<0.05$). Participants recall more agentive events than non-agentive events (Fig 2c; $z=4.4,\ p<0.05$).

In support of our framework, we find that linguistic and conceptual properties of events influence their recall, suggesting that encoding of semantic information reflects a utility function. By tailoring the utility function to a specific representation scheme, recall performance can be used as an evaluation metric for semantic representations. The varied nature of the properties investigated here suggest that models of semantic representation should be informed by representation schemes designed to explain many different behaviors (e.g., productivity, recognition and reasoning).

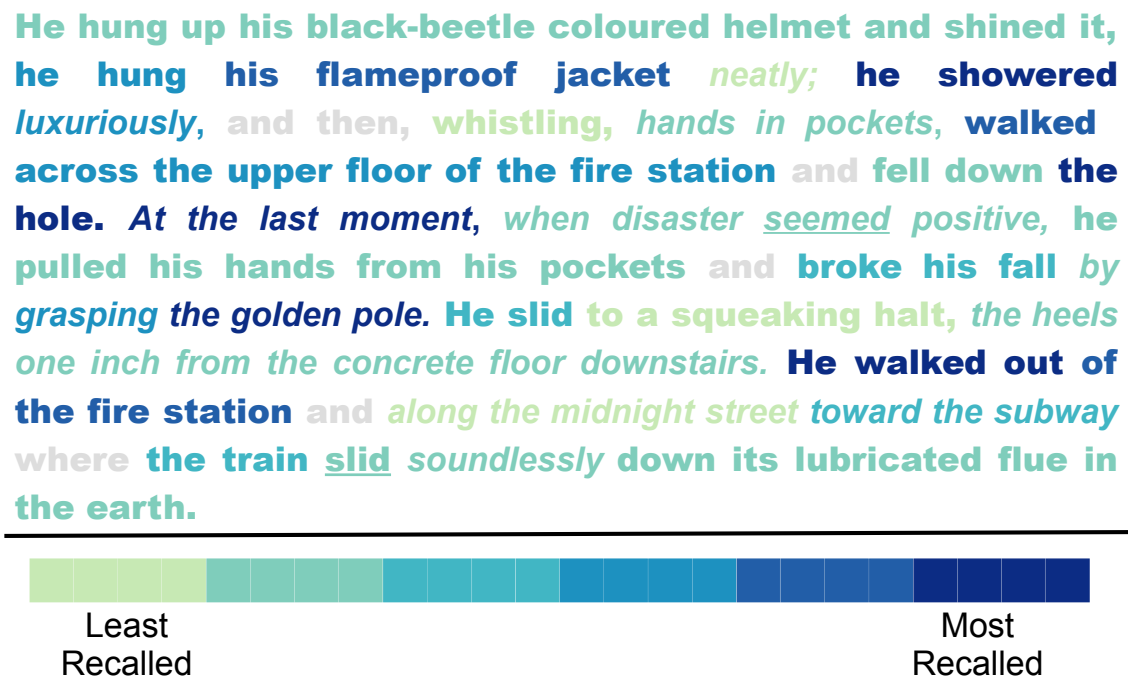

Figure 1: An example story read by participants. The text is shaded to reflect the proportion of participants that recalled those propositions. This is not exact as words may appear in multiple propositions. Adjuncts (in italics) are recalled less than arguments. Events without agents (underlined) are recalled less than events with agents.

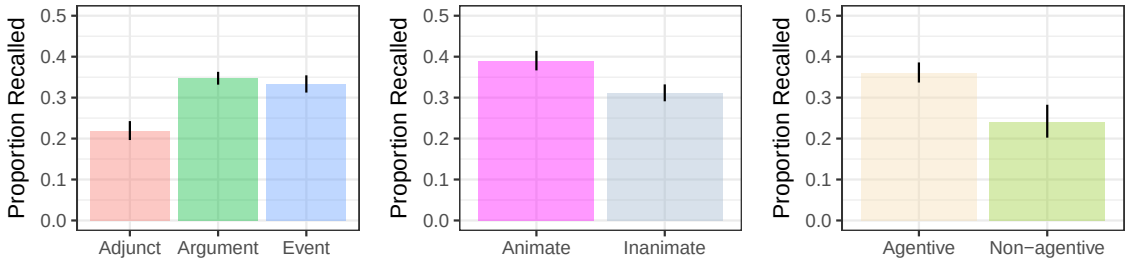

Figure 2: a) Proportion of arguments, adjuncts and events recalled. The equivalent proportions for arguments and events is expected given that events require their argument constituents. b) Proportion of animate/inanimate arguments recalled. c) Proportion of agentive/non-agentive events recalled. Line ranges reflect 95% binomial confidence intervals.

References

[1] McNamara, D. S., & Kintsch, W. (1996). Discourse processes, 22(3), 247. [2] Singer, M., & Ferreira, F. (1983). Journal of Verbal Learning and Verbal Behavior, 22(4), 437. [3] Kim, C. S., Gunlogson, C., Tanenhaus, M. K., & Runner, J. T. (2015). Cognition, 139, 28. [4] Kamp, H., & Reyle, U. (2013). Springer Science & Business Media. [5] Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). in The Conceptual Mind: New Directions in the Study of Concepts. MIT Press. [6] Collins, A. M., & Loftus, E. F. (1975). Psychological review, 82(6), 407. [7] Brown, P. M., & Dell, G. S. (1987). Cognitive Psychology, 19(4), 441. [8] Deese, J. (1959). Journal of experimental psychology, 58(1), 17. [9] Roediger, H. L., & McDermott, K. B. (1995). Journal of experimental psychology: Learning, Memory, and Cognition, 21(4), 803. [10] Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Journal of verbal learning and verbal behavior, 14(2), 196-214. [11] Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). Brain and language, 100(3), 223. [12] Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Journal of Experimental Psychology: General, 142(3), 880. [13] Bartlett, F. (1932). Cambridge: Cambridge University Press. [14] Bergman, E. T., & Roediger, H. L. (1999). Memory & Cognition, 27(6), 937-947. [15] Clarke, J., Srikumar, V., Sammons, M., & Roth, D. (2012). In LREC, 3276. [16] Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). Association for Computational Linguistics, 86.