

Event categorization using non-verb cues

Aron Marvel, Jean-Pierre Koenig & Gail Mauner (University at Buffalo)
aronmarv@buffalo.edu

Extant work on event categorization assumes verb senses denote event categories (e.g. McRae et al., 2005a; Vinson & Vigliocco, 2008), here defined as internal representations of bounded regions of space-time that can be characterized by features. But a single verb sense can denote several event categories. For example, an event where someone *raises* a glass in celebration (a toast) and an event where a crane *raises* a truck out of a lake are unlikely to be categorized together. While it has been recognized that particular non-verb cues generate expectations about upcoming content (see McRae et al., 2005b; Elman, 2009; Matsuki et al., 2011), it is unclear whether such cues are utilized for event categorization. The research reported on here addresses two questions: **(1) Do speakers use event participant properties to distinguish among event categories**, and **(2) if so, which of these properties matter most, both in general and for specific semantic domains?** To answer these questions, two experiments were conducted.

In Experiment 1, informed raters (the researchers) assigned categories to 2,000 pseudo-randomly selected sentences with one of 10 verbs from the British National Corpus, using parameters known to inform category judgments: agent type, time frame, event complexity, sociocultural salience, available inferences, and specific motion sequence. We presented 30 participants with 96 pairs of these sentences, balanced across 3 groups: (1) same verb sense + same rated category; (2) different verb sense + different rated category; and (3) same verb sense + different rated category. Participants judged the similarity of the events described by pairs of sentences on a 7-point Likert scale. A participant's overall median score was used as a breakpoint to determine whether pairs of sentences should be assigned to the "same" or "different" event category. Rater and participant judgments were compared in a Chi-square test of independence. The results strongly suggest a relationship between event similarity and inclusion of events within the same category, with a medium to large effect size as measured by Cramer's V ($X^2 = 218.64$, $N = 1129$, $p < 0.001$, $V = 0.44$), and further suggest that speakers are able to make category distinctions even when events share a verb sense (Figure 1). In Experiment 2, 120 participants each sorted into categories, based on the similarity of the events described, 6 sets of 20 sentences from the American National Corpus with the same verb and verb sense. Participants then listed the event properties they used to distinguish event categories when sorting. Verbs were balanced across 6 semantic domains (Table 1). Approximately 2,000 unique event features were elicited. Since many listed features could be near-synonyms, features were condensed into a smaller number of feature clusters through k-means clustering based on their semantic similarity (as per Latent Semantic Analysis) and a suite of optimality indices. These 'standardized' features were ranked according to their frequency of use and their distinctiveness as measured by cue validity. A series of mixed effects logit regressions suggested that 5 of 6 semantic domains significantly predicted an increased likelihood of specific standardized features being used for categorization (Table 1).

Experiment 1 suggests that speakers *do* use event participant properties to map the same verb senses into distinct event categories while Experiment 2 provides a broad set of non-verb features used for event categorization. Importantly, these data specify what kinds of non-verb information expanded theories of event categorization need to include, while still allowing verb senses to capture relevant information at a particular level of granularity. Furthermore, the importance of animacy, humanness, and plurality in participants' sorting suggests that the use of these properties in inflection and in agreement patterns across human languages might be rooted in their relevance to event categorization.

Elman, J. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science* 33(4), 547-582.

Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011;37(4):913.

McRae, K., Cree, G.S., Seidenberg, M.S., McNorgan, C. (2005a). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers* 37(4), 547-559.

McRae, K., Hare, M., Elman, J. & Ferretti, T. (2005b). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33(7): 1174–1184.

Vinson, D.P. & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavioral Research Methods* 40(1):183-90.

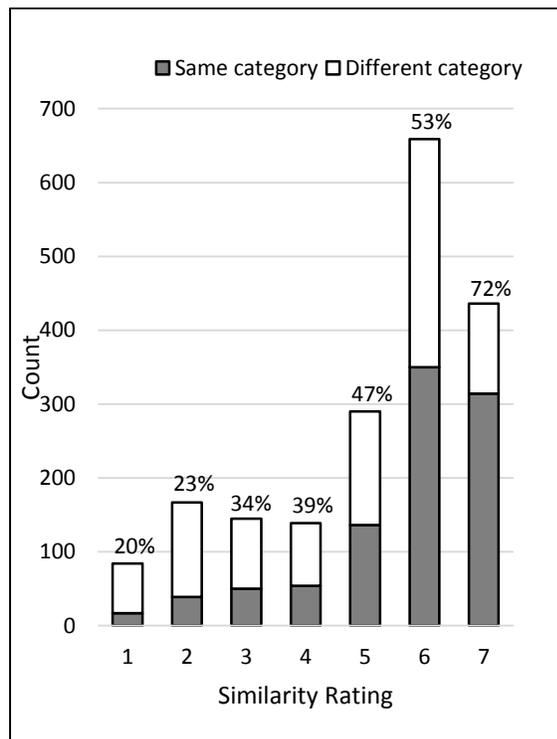


Figure 1. Percentage of same-category sentence pairs across similarity ratings (shared verb sense).

Semantic domain	Event participant features	Pr(> z)
Feeling	Group participants	< 0.001
	Government/military/politics	< 0.001
	Events involving people	< 0.001
Possession	Inanimate object participants	0.002
	Items or money	< 0.001
Movement	Animal participants	< 0.001
	Group participants	< 0.001
	Inanimate object participants	< 0.001
	Individual participants	0.01
Physical action	Animal participants	< 0.001
	Inanimate object participants	< 0.001
	Events involving people	< 0.001
	Individual participants	0.03
Mental	Group participants	< 0.001
	Events involving people	< 0.001
	Individual participants	< 0.001

Table 1. A verb's semantic domain predicts the increased use of specific properties of event participants for categorization.