MICROSENSES: EXPLORING MULTI-WORD EVENT CATEGORIZATION

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

One function of verbs is to name event categories. But event categories can receive their "names" from the combination of several words. An event category named by a single verb in one language can be conventionally coded via multi-word expressions in others (compare Spanish *entrar* and English *go in*). Furthermore, sentences with the same verb used in the same sense, according to best lexicographic practice, often belong to different event categories. Compare a) *The senator raised a glass in celebration* and b) *The crane raised the car out of the water*. While (a) describes a toast, (b) describes the extraction of a large object. Because the verb sense is unchanged, the information relevant for event categorization must come from sources outside of the verb and these sources combine with the verb to form complex event names.

Two questions motivate the research presented here. First, to what degree can native speakers recognize event categories that go beyond verb senses? Second, how can linguistic resources outside the verb serve as cues in making such distinctions? We explore answers to these questions in three ways. A rater study contrasts event categories with verb senses and gauges the relative contribution of subjects and objects as cues for categorization. A native speaker judgment task is then used to validate the results of the rater study and offers a more fine-grained representation of event similarity on a Likert scale. And lastly, subject and object meanings are tested as cues for event categorization in a computational model using Latent Semantic Indexing (LSI) and hierarchical clustering.

Based on previous research, we propose six parameters for distinguishing event categories beyond verb senses. They are listed in Table 1. These parameters were used as guidelines to sort into event categoriesa pseudo-random sample of 200 sentences for each of ten verbs *bake, borrow, buy, cover, deliver, frighten, immerse, pull, rescue, sell* from the British National Corpus (BNC). For simplicity, only verb plus subject and verb plus direct object combinations were analyzed. Results revealed an average 16.5 event categories per verb, in comparison to 3.8 senses according to the American Heritage Dictionary. 62% of event categories were distinguished by verb plus direct object combinations. This asymmetry was reversed for the object-experiencer verb *frighten*, suggesting the differential contribution comes from differences in semantic properties of the event participant types that typically occur in subject and direct object positions.

Results from the rater study were validated in an experiment where 30 participants were presented pairs of sentences from our corpus and asked to judge the similarity of the described events. Two of our ten verbs (*bake, immerse*) were excluded due to relatively sparse corpus data. The remaining eight verbs were used to construct 96 sentence pairs, balanced across three groups: in the first group each sentence pair shared both verb sense and event category according to our parameters; in the second group each pair differed in both sense and category but shared the same verb; in the third group each pair shared a verb sense but differed in category. The median similarity rating for each participant was used as a breakpoint and responses compared against same/different event category judgments from the rater study in a Chi-square test of independence. The results strongly suggest a relationship between similarity and inclusion of events within the same category, with 72% of similarity ratings in accordance with category judgments from the rater study and a medium to large effect size as measured by Cramer's V (X^2=218.64, N=1129, p<.001, V=.44). The results are shown in Figure 1.

The categorization and similarity judgment results suggest that event categories are distinguished at a more fine-grained level than that of the verb sense (at a ratio of at least 4:1), and that the contribution of participants named by subjects and objects in making that distinction is not symmetrical. But the results do not give us a way of quantifying the knowledge that *raising a car* is much more like *raising a truck* than either is like *raising a glass*. To quantify this knowledge, Latent Semantic Indexing (LSI) and a semantic space constructed from the BNC

were used to obtain relatedness values for subjects and direct objects. A mixed effects logit model suggested that LSI event participant relatedness values are reliable predictors of category inclusion, with higher relatedness values indicating greater odds of sharing an event category (S β=179.14, z=19.87, p<.001, DO β=69.87, z=12.03, p<.001). LSI relatedness values were also a marginally significant predictor of the odds of an item's being above a participant's median similarity rating (subject β=2.31, z=2.07, p=.04; direct object β=-1.70, z=1.58, p=.12).

The relationship between LSI relatedness and category inclusion allowed us to develop and test a computational model of event categorization. Pair-wise LSI relatedness values were obtained for each verb's subjects and separately its direct objects. Values were clustered using average linkage, and a cross-section of the clusters was taken when the number of clusters was equivalent to the number of the rater study event categories for that list. The resulting cluster categories were compared against the rater study categories, using the latter as a Gold Standard. An F value was calculated from the harmonic mean of the precision and recall measurements for the cluster categories. As a measure of comparison, each list was assigned 100 iterations of random categories and the resulting F value compared to the F obtained from LSI relatedness. On average, using LSI relatedness values gave a 42% improvement over random categories.

Research in semantics has shown event categories are not always named by single verbs. In this paper, we explore parameters for distinguishing event categories beyond single verb senses. Our results suggest (1) that raters can distinguish four times more event categories than dictionaries recognize verb senses, (2) some types of participants contribute more than others to distinctions among event categories and (3) similarities between event participants are good cues for distinguishing among event categories that go beyond single verb senses.

Agent type: Do the events involve agents of a distinct number, animacy, abstractness? E.g. *the fleet rescued* vs. *the medic rescued*
Time frame: Do the events last longer/shorter? E.g. *built a sand castle* vs. *built a skyscraper*
Sociocultural salience: Do the events play a special role in the practices of a community? E.g. *borrowed a book* vs. *borrowed an iron*
Available inferences: Is there additional pragmatic/semantic information obtained from world knowledge about typical event scenarios that the events take part in? E.g. *covered her bruise* vs. *covered her eyes*
Complexity: Do the events include particular subevents as well as relations among sub-events and their participants? E.g. *bought a house* vs. *bought sunscreen*
Specific motion sequence: Are actions performed according to a recognizable motor program? E.g. *pulled the ball* (soccer) vs. *pulled the cart*

Table 1*. Parameters relevant for event categorization*

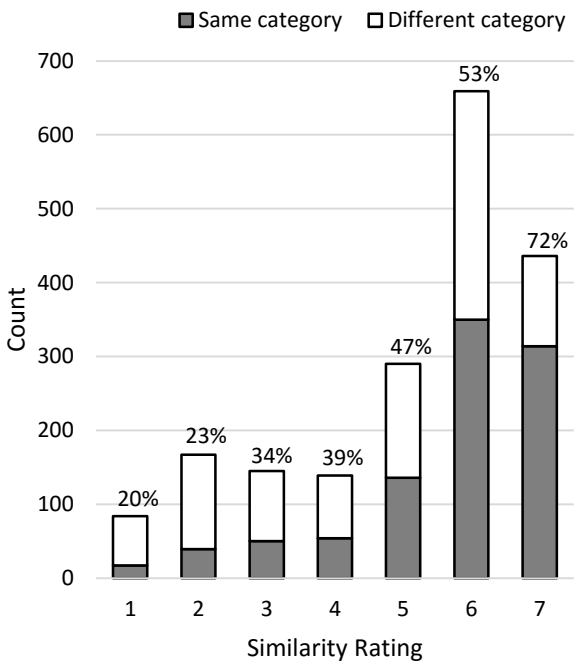

Figure 1. *Count and percentage in same category for each similarity rating (includes only pairs which share a verb sense)*