The Categorization of Intransitive Verbs: Evidence from word2vec modeling and a behavioral experiment

Kai-ying Lin (kylin@hawaii.edu), Peter Y. Washington (pyw@hawaii.edu)

University of Hawai'i at Mānoa

Introduction

- Unaccusativity Hypothesis (Permultter, 1978) suggests the subject of an unergative verb shows similar syntactic and semantic properties to the subject of an unaccusative verb, while the subject of an unaccusative verb shows similar properties to an object of an unaccusative verb.

  Unergative verbs (subject): runner, singer, swimmer, listener
  Transitive verbs(subject): eater, drinker, player, driver
  Unaccusative verb(subject): *dier, *comer, *disappearer, *arriver

- Projectional approaches (Chomsky, 1981; Hale & Keyser, 1993; Levin & Rappaport Havov, 1995; Pinker, 1989) emphasize that the **meaning** of the verb plays a more definitive role in deciding the category of the verb.
- Constructional approaches (Hoekstra, 1992; Borer, 1994; Goldberg, 1995; van Hout, 1996; Ritter & Rosen, 1996) stress the **constructional environment** in which the verb appears can decide the category of the verb.
- Word2vec serves as an efficient tool to examine both approaches as it generates semantic embeddings using the association of surrounding words.
- A child acquisition experiment provided with semantics and constructional environment of new verbs can examine the categorization and demonstrate both approaches.

Word2vec modeling

- Constructional environment and frequency of a new verb occurring in constructional environment were examined.
- Constructional environment: 11 types of pre-selected sentences were filled with 22 new verbs (one type of pre-selected sentence has 2 new verbs).
- Frequency: the occurrence of each verb (and each sentence) was controlled to 4, 8, 32 and this yields 88, 176, 704 pre-selected sentences in total.
- Word2vec model was trained with Taiwanese Mandarin CHILDES data, as well as 11 types of pre-selected sentences.
- Each pre-selected sentence was assigned a category of Unaccusativity based on research outcomes in the literature, as shown below.

Sentence No.	Pre-selected sentence (constructional environment)	Expected effects of categories
1	LVS	unaccusative
2	Resultative (2nd)	unaccusative
3	Ba-sentence	unaccusative
4	Perf (null-subject)	unaccusative/neutral
5	Perf	unaccusative/neutral
6	Resultative (1st)	neutral
7	Resultative(1st-null)	neutral
8	Will+V	neutral/unergative
9	V+Dur	unergative
10	Want_V	unergative
11	Imperf+V	unergative

Results

- The result met the expectation in the table.
- The cosine similarities of new verbs are compared with existing unergative and unaccusative verbs in the CHILDES corpus to show results of categorization.
- The cosine similarities of new verbs show clustering when occurring in different constructions ($p<.001$**) and more occurrences of sentences can contribute to more obvious clustering ($p<.001$***).
- The constructional effect can be more than 2 or 3 categories, forming gradience.

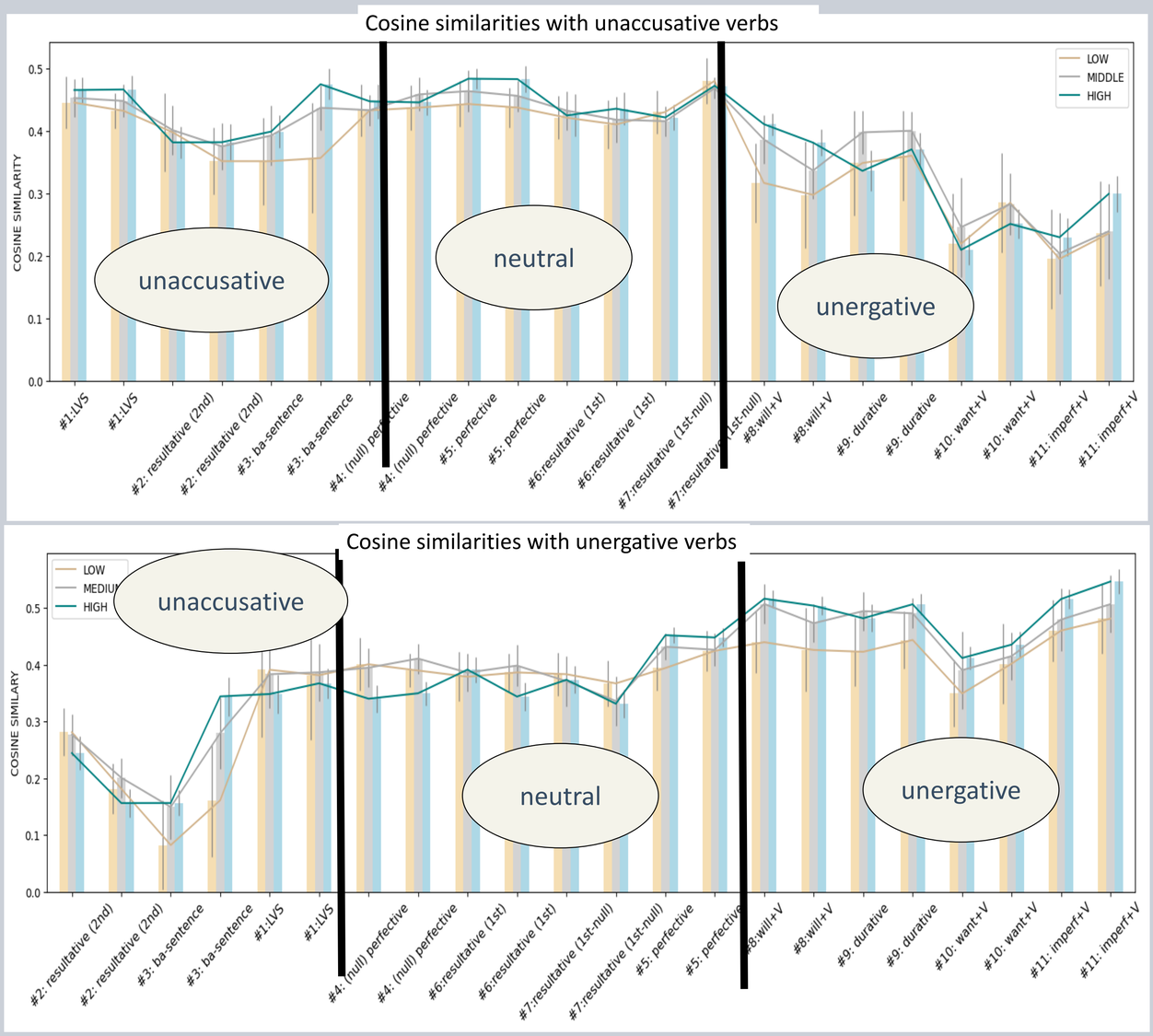

Figure 1, 2: Average cosine similarities of new verbs compared to existing unaccusative/unergative verbs. LOW, MIDDLE and HIGH models have varying frequency of pre-selected sentences and new verbs. The Figure is divided into three partitions, each representing a sentential category as shown in the box.

Behavioral experiment

- 66 Mandarin-speaking children (4;10 to 6;4, mean=5;6) were divided into a non-construction and a construction group.
- 6 new verbs were created and put into the constructional environment of #1, #2, #1 and 2, #11 as listed above.
- The semantics of new verbs targeted the difference of having the endpoint or not during the event (telicity) and were represented as animations.
- The non-construction group received only the semantics of new verbs, while the construction group received the semantics of new verbs within the constructional environment.
- Children were asked to rate diagnostic sentences to show their results of categorization.
- Ratings were converted into accuracy when they aligned with the expected categories.
- The 'accuracy' showed that all the constructional environments can aid in the categorization and group is a significant predictor ($p<.001$***).

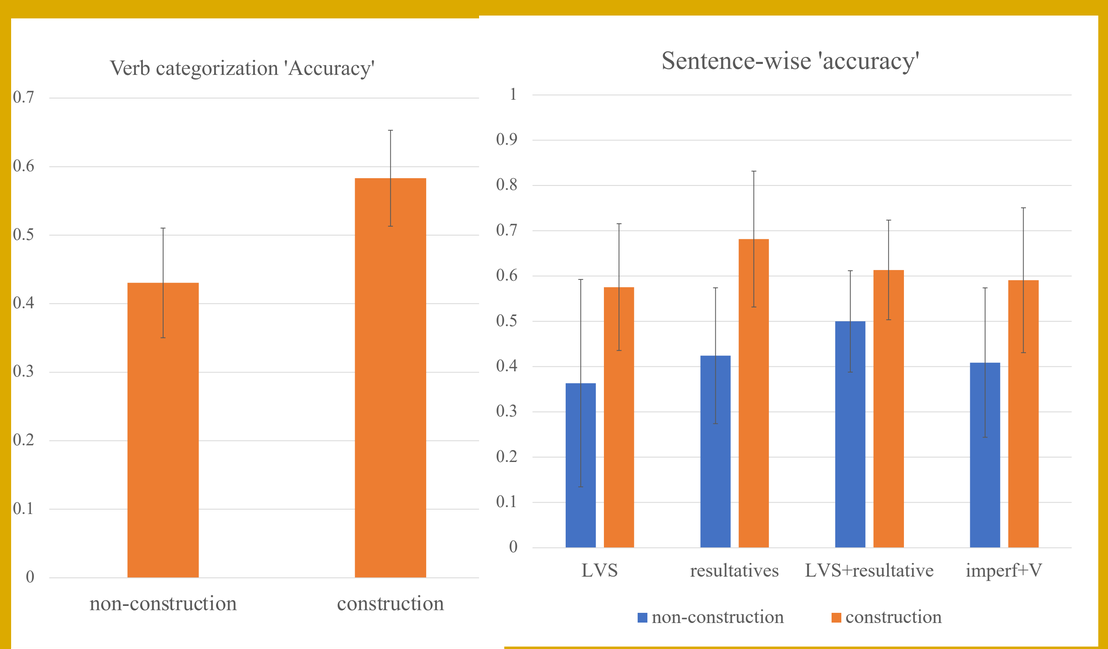

Figure 3, 4: group-wise accuracy and sentence-wise accuracy

Selected References

Perlmutter, D. (1978). Impersonal passives and the unaccusative hypothesis. In J. Jaeger (Eds.), *Proceedings of the 4th annual meeting of the Berkeley Linguistics Society* (pp. 157-190). Berkeley, CA: Berkeley Linguistics Society. Borer, H. (1994). The projection of arguments. In Benedicto, E., & Runner, J. (Eds.), *University of Massachusetts occasional papers in Linguistics 17*, University of Massachusetts. Chomsky, N. (1981). *Lectures on government and binding*. Hawthorne, NY: Foris. Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press. Hoekstra, T. (1992). *Aspect and theta-theory*. In Roca, I. M. (Ed.), *Thematic structure: Its role in grammar* (pp. 145-174). Berlin, Germany: Mouton de Gruyter. Hale, K., & Keyser, S. J. (1993). On the argument structure and the lexical expression of Syntactic Relations. In Hale, K. & S. J. Keyser. (Eds.), *The view from building essays in honor of Sylvain Bromberger* (pp. 53-109). Cambridge, MA: MIT Press. van Hout, A. (1996). *Event semantics of verb frame alternations. A case study of Dutch and its acquisition*. Tilburg, Netherlands: University of Tilburg. Levin, B., & Rappaport-Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT Press. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. In Butt, M., & Geuder, W. (Eds.), *The Projection of arguments: Lexical and compositional factors*, 97-134. Stanford, CA: CSLI Publications.

Conclusion

- Projectionists' and constructionists' considerations are cognitively universal
- The constructional environment and its occurrences can influence the categorization of intransitive verbs
- The semantics of the verbs also plays a role in the categorization
- The constructional effect can be gradient, more than the expected three categories