

## Form, Function and Productivity of English Light Verb Constructions

Consider for a moment the simple, everyday English phrases, *Sarah took a bath* and *Sarah gave the baby a bath*. What events are denoted by these phrases? If you were not a proficient English speaker, you might look up the words in the sentence. The first definition for *take* appearing in the American Heritage Dictionary (<http://www.ahdictionary.com>) is “to get into one’s possession...” The first definition for *give* is “to make a present of.” Thus, one could logically conclude that *giving* and *taking* events involve the transfer of possessions. What object is being transferred here? *A bath*: “a vessel for holding water in which to wash the body.” Yet, for fluent speakers of English, these phrases do not conjure an image wherein Sarah is picking up a bath tub and either taking it somewhere or giving it to a baby; rather, both phrases refer to events where Sarah herself bathed or bathed the baby.

These phrases are often referred to as “Light Verb Constructions” (LVCs) (Jespersen, 1942); however, both the labeling and definition of such constructions remain under debate in the linguistic community. LVCs also include expressions like *have a drink*, *make an offer*, *take a walk*, and *do an investigation*. LVCs consist of a semantically general verb and a noun that denotes an event or state.

Dominant theories in linguistics, such as the Chomskyan or “generative” approach, would predict an interpretation of the sentence that was based on the meanings of the component words alone. Specifically, generative syntax assumes that syntactic rules themselves, which predict how the elements of the language like nouns and verbs can grammatically combine, do not carry any semantics. Rather, semantic facts arise entirely from the lexicon (Chomsky, 1965). LVCs do not seem to fit with this picture, because syntactically, these constructions are identical to a typical *verb + noun* combination in which the verb is the main predicate and the noun is an argument of the verb (e.g. *Sarah took an apple*). Yet speakers interpret LVCs differently from typical *verb + noun* combinations and are able to recognize that the event or state being described is expressed by the noun. Thus, LVCs present evidence that particular combinations of words carry special meaning; meaning that goes beyond the sum of the component parts, an idea that has been recognized by Construction Grammar (see for example, Goldberg, 1996, 2006; Croft & Cruse, 2004).

The importance of LVCs has become clear in Natural Language Processing (NLP). The successes of supervised approaches to automatic semantic analysis (e.g. Márquez et al., 2008) rely on quality lexical resources such as PropBank (Palmer et al., 2005), VerbNet (Kipper et al., 2008), and FrameNet (Baker et al., 1998; Fillmore et al., 2002), which provide information on the relational semantics of verbs. However, automatic systems are limited by the static nature of these resources, which do not reflect the potential for speakers to use a verb in new usages like LVCs. LVCs are especially problematic for such resources because they are “semi-productive,” meaning that feasibly new LVCs can enter the language (Nickel, 1978: 83); for example, *make a backup* in the technology domain. However, the productivity is not unlimited, and determining how and why it is limited to certain constructions is a difficult question. For example, while *take a walk* and *take a stroll* are acceptable to most speakers, *take a run* is not. To supplement lexical resources with LVCs, it is necessary to define and understand the function and productivity of LVCs. Improved lexical resources may in turn help NLP systems to correctly identify and interpret LVCs.

ICS researchers are combining machine learning approaches, corpus studies, and human judgment tasks to better understand LVCs. This research, funded by the National Science Foundation with principle investigator Martha Palmer, uses machine learning techniques to automatically detect LVCs, corpus studies to understand their function in context, and human judgment tasks to understand both their interpretation and productivity. This research will therefore provide important insights into competing linguistic theories, and will help to improve existing NLP lexical resources.

## References

- Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*. Part VI: *Morphology*. With assistance of P. Christopherson, N. Haislund, & K. Schibbye. London: George Allen & Unwin. Copenhagen: Ejnar Munksgaard.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press.
- Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck. 2002. Background to FrameNet. *International Journal of Lexicography*, 16(3):235-250.
- Goldberg, Adele. 1996. Construction Grammar. In Keith Brown and Jim Miller ( eds.), *Concise Encyclopedia of Syntactic Theories*. New York: Elsevier, 68–71.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. *Proceedings of the 17th International Conference on Computational Linguistics (COLING/ACL-98)*, pp. 86–90, Montreal.
- Croft, William and D.A. Cruse. 2004. From Idioms to Construction Grammar. Chapter 9, *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation Journal*, 42:21–40.
- Márquez, L., X. Carreras, K. Litkowski, and S. Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2): 145-159.
- Nickel, Gerhard. 1978. Complex Verbal Structures in English. *Studies in Descriptive English Grammar*; 63-83.
- Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.