Automatic identification of motion constructions in L1/L2 English spoken corpora

Hakyung Sung, Department of Linguistics, University of Oregon, hsung@uoregon.edu
Kristopher Kyle, Department of Linguistics, University of Oregon, kkyle2@uoregon.edu

An argument structure construction (ASC) can be defined as a clausal form that corresponds to a propositional meaning (e.g., Diessel, 2004; Goldberg, 1995). It has been an important area of investigation to analyze the relationship between ASC use and language acquisition/development in both first (L1) and second (L2) language learning research (e.g., Diessel, 2013; Ellis, 2002). In this paper, we focus on English motion constructions, which include intransitive motion constructions (e.g., *the fly buzzed into the room*) and caused-motion constructions (e.g., *Tom kicked the ball into the net*). The acquisition of motion constructions in L1 has often been discussed in relation to the cognitive development of spatial and motional concepts in human experience (Goldberg, 1995, 2006; Talmy, 1985), while most of the L2 English studies emphasized typological differences and L1 influence on the acquisition of the L2 motion constructions. While the researchers in this field have widely used L1/L2 corpora to track the acquisition/development patterns in learners' production data, previous research has primarily been limited to small datasets and time-consuming manual annotation (e.g., Ellis & Ferreira-Junior, 2009a, b).

In this study, we empirically evaluate the performance of a state-of-art multi-class transformer-based ASC annotation model (Kyle & Sung, 2023) when used to annotate motion constructions in L1 and L2 spoken corpora. The L1 data consisted of sentences from responses to picture description tasks from two corpora [ALLSSTAR (Bradlow et al. 2010); ICNALE SD (Ishikawa, 2019)]. These data represent responses from 46 participants. The L2 data consisted of sentences from responses to a narrative retelling task produced by Korean L2 English speakers (n=94). The researchers manually annotated intransitive-motion constructions (n = 375), caused-motion constructions (n = 628), and other constructions (n = 1030), and then also automatically tagged the sentences using the ASC annotation tool. F1 scores were calculated by comparing the manually and automatically annotated versions of each sentence.

Overall, the results indicated relatively high tagging accuracy for both datasets (Table 1). The lowest accuracy scores were found for caused-motion constructions in the L1 dataset (F1 = 0.84). These findings suggest that the automatic ASC tagger can be used to annotate L2 spoken data collected from narrative retelling and picture description tasks with a relatively high degree of confidence. The presentation will discuss the strengths and weaknesses of the ASC annotation tool, with a focus on accuracy for particular verbs and ASCs, and future directions for improvement of the performance of the tool based on the study's findings.

	Freq	Precision	Recall	F1
L1 caused-motion construction	166	1.00	0.72	0.84
L1 intransitive motion construction	137	0.96	0.94	0.95
L2 caused-motion construction	462	0.99	0.90	0.94
L2 intransitive motion construction	238	0.95	0.93	0.94

Table 1. F1 scores

References

Bradlow, A. R., Ackerman, L., Burchfield, L. A., Hesterberg, L., Luque, J. S., & Mok, K. (2010). ALLSSTAR: Archive of L1 and L2 scripted and spontaneous transcripts and recordings. In *Proceedings of the International Congress on Phonetic Sciences,* 356-359.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge University Press.

Diessel, H. (2013). Construction grammar and first language acquisition. *The Oxford handbook of construction grammar*, *347*, 364.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143-188.

Ellis, N. C., & Ferreira-Junior, F. (2009a). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, *7*(1), 188-221.

Ellis, N. C., & Ferreira–Junior, F. (2009b). Construction learning as a function of frequency, frequency distribution, and function. *The Modern language journal*, *93*(3), 370-385.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press on Demand.

Ishikawa, S. I. (2019). The ICNALE Spoken Dialogue: A New Dataset for the Study of Asian Learners' Performance in L2 English Interviews. *English Teaching*, *74*(4), 153-177.

Kyle, K., & Sung, H. (2023). An Argument Structure Construction Treebank. In *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+ NLP, GURT/SyntaxFest 2023)* (pp. 51-62).

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 3, pp. 71-149). Cambridge: Cambridge University Press.