

A Best-Fit Approach to Productive Elision of Arguments

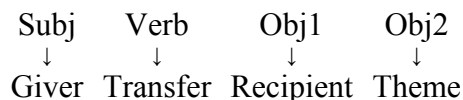
Languages such as Chinese allow productive, non-idiomatic argument elision even in written text^[1, 2]. Most construction-based theories require enumeration of all possible argument combinations within the grammar. We show that a best-fit sentence analysis model situated within a language understanding framework simplifies the representation of argument structure constructions by flexibly allowing elision depending on each construction and its context, thereby obviating the need for explicit enumeration of argument combinations in the grammar. In essence, we leverage process instead of representation to model elision.

The best-fit analysis framework is a contextually and semantically driven process consistent with the competition-based processing models discussed in the psycholinguistic literature^[3, 4]. Recalling that constructions are pairings of form and meaning, the best-fit model finds the most-likely analysis of a sentence given the context and the particular form and meaning bindings that the analysis entails.

Examples 1-6 show productive, non-idiomatic argument dropping in the ditransitive construction in Mandarin Chinese. These are taken from the Tardiff Beijing Corpus^[5, 6] in CHILDES^[7], a corpus of caregiver-child interaction. Example (1) exhibits the construction without any argument dropping – it has a verb of transfer, a subject (the giver) and two objects (the recipient and the theme). Of the 51 phrases containing the verb *gei3* (*give*) in two transcripts from the same parent, only three have all the arguments present. All other argument dropping patterns are attested, with five shown below. Examples like (2), (3) and (4) with one argument dropped account for almost half the data, and examples like (6) where all arguments are dropped are observed on seven occasions, not counting repetitions of a preceding child utterance.

- | | |
|--|---|
| <p>1 ma1+ma gei3 ni3 zhei4+ge .
mother give 2PS this+CLS
<i>Mother (I) give you this (a toy).</i></p> | <p>2 ni3 gei3 na2 yi2+ge da4 kuair4 de .
2PS give that 1+CLS big CLS NOM
<i>You give that big one (a peach) [to auntie].</i></p> |
| <p>3 ni3 gei3 yi2 .
2PS give auntie
<i>You give auntie [the peach].</i></p> | <p>4 gei3 a1+yi2 yi2+ge qu4 ao .
give auntie 1+CLS DEIT_{away} EMP
<i>[You] Give auntie one (peach).</i></p> |
| <p>5 ao ni3 gei3 ya .
EMP 2PS give EMP
<i>Oh you give [auntie] [the peach].</i></p> | <p>6 gei3 .
give
<i>[I] give [you] [some peach].</i></p> |

The best-fit model handles the above variations in argument elision with just one ditransitive construction, shown below. Crucially, for each constituent in the construction, the model probabilistically allows for it to be local or elided, maximizing the constructional, semantic, and contextual fit to the sentence. An existing sentence analysis model^[8] is capable of finding the constructional and semantic best-fit analysis using Embodied Construction Grammar (ECG)^[9], and this work is currently being extended to take into account the situational and discourse context so that elided arguments can be recovered. The proposed technique is being applied in language understanding and in a model of how children learn a language like Mandarin that elides many arguments.



References:

- [1] C.-L. Yeh and Y.-C. Chen. (2004). Zero anaphora resolution in Chinese with shallow parsing. Accepted for publication by *Journal of Chinese Language and Computing*.
- [2] W. Li. (2004). Topic chains in Chinese discourse. *Discourse Processes*, Vol. 27, No. 1, 25-45.
- [3] J. C. Sedivy. (2002). Invoking Discourse-Based Contrast Sets and Resolving Syntactic Ambiguities. *Journal of Memory and Language* 46, 341–370.
- [4] S. Narayanan and D. Jurafsky. (2002). A Bayesian Model Predicts Human Parse Preference and Reading Times in Sentence Processing. *Advances in Neural Information Processing Systems 14*. Cambridge, MA: MIT Press. 59-65.
- [5] T. Tardif. (1993). Adult-to-child speech and language acquisition in Mandarin Chinese. Unpublished doctoral dissertation, Yale University.
- [6] T. Tardif. (1996). Nouns are not always learned before verbs: Evidence from mandarin speakers' early vocabularies. *Developmental Psychology*, 32, 492–504.
- [7] B. MacWhinney. (2000). *The CHILDES project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- [8] J. Bryant. (2003). *Constructional Analysis*. Master's Thesis, University of California, Berkeley.
- [9] B. K. Bergen and N. Chang. (2005). Embodied Construction Grammar in simulation-based language understanding. In Jan-Ola Östman and Mirjam Fried, eds. *Construction Grammars: Cognitive Groundings and Theoretical Extensions*. Philadelphia, PA: John Benjamins. 147-190.