

Verb Subcategorization in Indonesian

Fillmore (1970) in his seminal paper “*The Grammar of Hitting and Breaking*” recognized that verbs in English could be grouped into classes based on their semantic similarity as well as shared grammatical behavior and argument realization. Specifically, he showed how *hit* and *break* verbs are each members of larger classes of verbs whose members share comparable patterns of behavior such as participation in the causative alternation and interpretations available to their passive particles (1970:125). Other studies have since been done in English which confirmed and expanded on Fillmore’s findings (Dixon 1991; Jackendoff 1990; Levin & Rappaport Hovav 1991), most notably by Levin (1993) in her book “*English Verb Classes and Alternations*”. Moving beyond English, the idea of semantically related verb classes having shared syntactic behaviors has also been identified and explored in other languages such as Lhasa Tibetan (DeLancey 1995), Kimaragang Dusun (Kroeger 2010), Jarawara (Vogel 2005) and Indonesian (Voskuil 1995). Most recently, this idea has been implemented computationally in Hebrew by Scheinfx et. al (2016). Their study proposed an analysis that explained argument structure phenomena in Hebrew by distinguishing between semantic and syntactic selection and stating the constraints in each level separately.

This study attempts to capture the semantic and syntactic generalizations of verbs in Indonesian, more specifically for an existing subclass of verbs that take the suffixes *-i* and *-kan* (Anuar and Choi, 2017). A section of the Indonesian sub-corpus found in the the Leipzig Corpora Collection (LCC Quasthoff et al. 2006) containing over 15.5 million sentences from news and web articles was searched for surface forms prefixed with *meN-* and *di-*, optionally suffixed with *-i* and *-kan*. 4838 roots with 116692 different surface forms were selected for further subcategorization, with four major syntactic groups:

	Group 1	Group 2	Group 3	Group 4
	+kan, +i	+kan, -i	-kan, +i	-kan, -i
<i>meN-</i> or <i>di-</i> prefixed roots	<i>menempat</i> <i>menempatkan</i> <i>menempati</i> ‘to place’	<i>menghilang</i> <i>*menghilangi</i> <i>menghilangkan,</i> ‘to disappear’	<i>melompat</i> <i>melompati</i> <i>*melompatkan</i> ‘to be located’	<i>mengacak</i> <i>*mengacakkan</i> <i>*menacaki</i> ‘to disorganize’
No. of roots	447	2089	385	1917
Total roots	4838			

Table 1: Number of roots found for the major syntactic groups of verbs in Indonesian

Out of these, 50 roots will selected for more in-depth analysis. 100 sentences for each root will be annotated for valency and semantic roles of their arguments, using a modified list of 10 semantic roles defined in Scheinfx et al. (2016). The annotated verbs will be analyzed and categorized into semantic classes according to their shared behaviors.

Bibliography

- Anuar, A and Choi, H. 2017. Verb Subcategorization: *-kan* and *-i* suffixing verbs in Malay and Indonesian. Paper presented at ISMIL 21.
- DeLancey, Scott. 1995. Verbal case frames in English and Tibetan. Ms., Department of Linguistics, University of Oregon, Eugene, OR.
- Dixon, R. M. W. 1991. *A New Approach to English Grammar, On Semantic Principles*. Oxford: Oxford University Press.
- Fillmore, C. J. 1970. The grammar of *hitting* and *breaking*. In: Jacobs, Roderick A. & Rosenbaum, Peter S. (eds.), *Readings in English Transformational Grammar*, 120–133. Waltham, MA: Ginn.
- Jackendoff, R. S. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.
- Kroeger, P. 2010. The grammar of *hitting* and *breaking* (and *cutting*) in Kimaragang Dusun. *Oceanic Linguistics* 49: 2–20.
- Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.
- Levin, B. 2013. Verb classes within and across languages. Stanford University.
- Sheinfx, L., Melnik, N., & Wintner, S. (2016). Representing argument structure. *Journal Of Linguistics*, 1-50. Cambridge University
- Quasthoff, U., Richter, M. and Biemann, C. 2006. Corpus portal for search in monolingual corpora. In Proceedings of the Fifth International Conference on Language Resources and Evaluation, 1799–1802. Genoa.
- Sneddon, J. N. 1996. *Indonesian: A Comprehensive Grammar*. London and New York: Routledge.
- Vogel, A. R. 2005. Jarawara verb classes. Ph.D. dissertation, University of Pittsburgh, Pittsburgh, PA.
- Voskuil, J. 1996. *Verb Taxonomy in Indonesian, Tagalog and Dutch*. Netherlands: HIL Dissertation.