

Discourse conditions on null anaphora in colloquial Indonesian

T. J. Conners, C. M. Brugman, and N. B. Adams

University of Maryland

Indonesian is quintessentially a “discourse-based pro-drop” language, in exhibiting a free willingness to leave out multiple arguments and adjuncts to predicates in speech. We report on a corpus-based study that relies on a corpus¹ of naturalistic spoken Jakarta Indonesian annotated by native speakers, including the annotation of zeroes, to discern what the discourse conditions on zeroes really are.

Whereas syntactic theories try to parameterize minimum conditions on the occurrence of null anaphoric elements with respect to clauses, these theories fail to account for the rampant occurrence of multiple nulls in clauses, not to mention admitting of discourse- or information-based conditions. These clause-based accounts fail when faced with a language that has no morphological agreement or agglutinative morphology, that does exhibit both thematic and generic null subjects, and that allows for *pro* to be controlled outside of its domain--all features of specific languages that have been successively adduced in articulating clause based accounts. Indonesian has all of these features; therefore, clause-based analyses provide no insight. The current paper therefore takes a corpus-based approach and applies the predictions of two discourse-based theories, namely Centering Theory and Accessibility Theory, to account for the attested distribution of null arguments in conversation. We limit ourselves to null subjects to avoid the problematic question of how to determine the transitivity of predicates in Indonesian.

Accessibility Theory (Ariel 1988, inter alia) holds that there is a hierarchy of types of referring expressions (based on the lexico-syntactic properties of the element, e.g. definite noun phrase, indefinite noun phrase, etc.) and their availability as potential antecedent to an anaphoric element—the latter implicating also lexico-syntactic properties of the anaphoric element, as well as discourse-based relations between them. Centering theory (Grosz et al., 1983, inter alia), while similar in some respects to Accessibility Theory, identifies some different elements of the items involved in a potential antecedency relation. From the Jakarta Indonesian data we will describe the effects of information (“given-new”) status; animacy/humanness; morphosyntactically marked Focus (*-lah/dong/demonstrative* marking) ; morphosyntactically marked Topic (*yang*-constructions). We will test the predictions made by Accessibility Theory and/or Centering theory with respect to each of these markers of “salience” or availability of referent.

A subset of the corpus was annotated by two speakers, to yield a somewhat surprising inter-annotator outcome which is also reported on as part of this study.

References

Ariel, M. 1988. Referring and Accessibility. *Journal of Linguistics* 24;65-87.

¹ Corpus from Jakarta Field Station, Max-Planck Institute for Evolutionary Anthropology, *Corpus of Jakartan Indonesian*.

Grosz, B., A. Joshi, and S. Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proc. 21st Annual meeting of the ACL*: 44-50. Association for Computational Linguistics.