

Structural and functional mismatches in Indonesian *ber*-constructions: issues in linguistic analysis and computational implementation

I Wayan Arka (ANU), Avery Andrews (ANU),
Mary Dalrymple (Oxford University), Jane Simpson (Sydney University)

Abstract

This paper reports the challenges we face in developing an Indonesian electronic grammar, particularly in handling structural and functional mismatches of *ber*-verb constructions exemplified in (1)-(2). The intransitive *ber*-verb, often morphologically complex, can be followed by an adjunct, i.e. *hitam* 'black' (1) or a complement, *orang Indonesia* 'Indonesian person' (2). Crucially, the adjunct or complement is associated with the noun stem inside the verb, as indicated by the arrow and seen from the translation (i.e. *hitam* modifies *sepatu* and *orang Indonesia* modifies *istri*). This structure represents a mismatch between structural and semantic/functional units. On the surface constituency, the adjunct or complement does not form a phrasal or clausal structural unit with the noun it modifies. The noun is a stem inside the *ber*-verb, yet at the semantic/functional level the adjunct or complement must be associated with the noun stem as a unit. Our linguistic analysis must capture this mismatch. We argue that this can be nicely modelled in Lexical Functional Grammar (LFG), which separates out surface constituency structure (*c-str*) from meaning representation (functional structure, *f-str*).

- (1) Ia *ber-sepatu* (*hitam*).
3s ber-shoe black
'S/he is wearing black shoes.'
'S/he is wearing shoes which are black.'
- (2) a. Ia *ber-istri-kan* *orang Indonesia*. b.* Ia *beristrikan*.
3s ber-wife-APPL person name
'He is having/has an Indonesian wife.'

Note that the *ber*-verb provides empirical evidence for word-internal elements being visible to syntax, and this must also be captured by the linguistic analysis. We adopt a version of lexical integrity principle (Bresnan and Mchombo 1995; Bresnan 2001) which allows elements within words to be accessible from certain word-external syntactic processes. We propose that the adjunct or complement be analysed as a predicative unit which takes the noun stem inside the *ber*-verb as its argument. We argue that the (classic) LFG model provides us a means to grab the stem inside the verb for the purpose of word-external modification required in this construction.

We demonstrate the progress that we made in our project in implementing and testing the analysis within the LFG-based XLE (Xerox Linguistic Environment) system. This includes how *ber*-sentences are tokenized and how the syntactically relevant information for *ber*- is collected and processed by the morphological analyser, which is then fed into the grammar to provide the correct output of *c-str* and *f-str* parsing of the sentence.

References

- Bresnan, Joan. 2001. *Lexical functional syntax*. London: Blackwell.
Bresnan, Joan, and Sam Mchombo. 1995. *The lexical integrity principle: evidence from Bantu*. *NLLT* 13:181-252.