A COMPUTATIONAL MORPHOSYNTACTIC ANALYSIS FOR THE APPLICATIVE *-I* IN INDONESIAN

Wayan Arka♦, Mary Dalrymple♣, Meladel Mistica♦, Avery Andrews♦, and Jane Simpson♥
♦ Australian National University, ♣ Oxford University, ♥ The University of Sydney

ABSTRACT

Indonesian is one of the most extensively studied Austronesian languages (Musgrave 2001, Chung 1976, Myhill 1988, Purwo 1989, Sneddon 1996, Macdonald and Dardjowidjojo 2001, among others), and has rich valence-changing morphology. This research examines the applicative and causative usage of the suffix *-i*, and its complex interaction with voice selection, as seen in example (1b) where the verb exhibits the applicative *–i* with actor voice (AV) marking. In addition to examining the implementation issues of our analysis, we also raise the sociolinguistic issues associated with the usage of Indonesian applicatives, *-i* vs. *–in* (formal vs. informal/colloquial registers).

(1). a. Ia duduk di kursi itu
3s sit LOC chair that
'S/he sat on the chair'

b. Ia menduduki kursi itu
3s AV.sit-*i* chair that
'She was sitting on the chair'

c. *Ia duduk kursi itu

Implementing an analysis of the suffix *–i* in a computational grammar requires addressing at least two key issues: i) how the suffix *–i* is morphologically identified as part of a larger system of verb formation in Indonesian, and ii) how to capture syntactic, semantic and aspectual information of *–i* so as to produce correct parses of sentences headed by verbs with *-i*. This paper reports the progress we have made on these two key issues in our Indonesian computational grammar. In implementing the LFG analysis of the applicative function of *–i* in XLE, we analyse *-i* as a verbaliser/transitiviser carrying its own PRED(ICATE) argument structure. Word-formation with *–i* involves predicate composition of the PRED of the suffix with the PRED of its stem, modelling complex predicate formation as described in Alsina (1996) and Butt (1995). Predicate composition in applicativisation (and its complex interaction with other kinds of word-formation such as voice selection and reduplication) has not been previously implemented in XLE. We implemented it making use of the restriction operator (Kaplan and Wekind 1993, Butt and King 2006, Butt, King and Maxwell III 2003), as shown in (2).

(2) APPLICATIVE *-i*

```
APPL_I( _P) =
        {(^ PRED) = '_P<(^ SUBJ)(^ OBJ) %PRED3>'
        ^\PRED\SUBJ\OBJ\OBL = !\PRED\SUBJ\OBJ\OBL
                {~(! OBJ)
                (! SUBJ) = (^ SUBJ)
                | (! OBJ)
                (! SUBJ) = (^ SUBJ)
                (! OBJ) = (^ OBJ)
                (^ TNS-ASP PROG)=+
                }
        }
        (! PRED)=(^ PRED ARG3)
        (^ APPLICATIVE)= +.
```

(3) TAGS AND THEIR ENTRIES

```
a. +I       V_I       XLE.
b. AV+      V_VOICE   XLE  @(VOICE-TYPE AV).
c. UV+      V_VOICE   XLE  @(VOICE-TYPE UV).
d. PASSdi+  V_VOICE   XLE @(VOICE-TYPE PASSIVE).
```

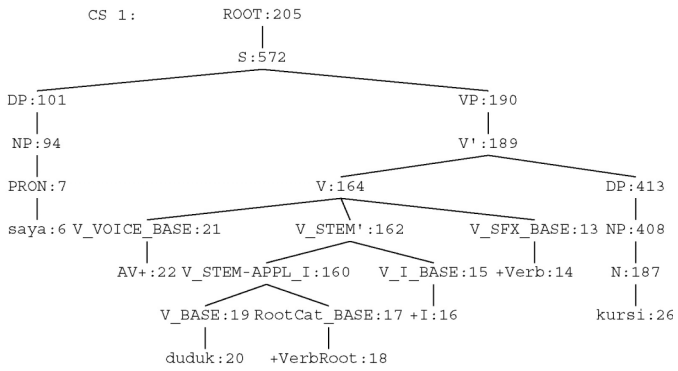
The identification of the morpheme *-i* is done by the tokenizer and morphological analyser which we have developed using XFST technology (Beesley and Kartuneen 2003). The input is tokenised, and the morphological analyser identifies the tokens as verbs and segments them into their morphological units. The morphological analyser contains morphophonemic and morphotactic rules that enable it to recognise different kinds of affixation including *-i*. Our current version of the morphological analyser is capable of analysing complex word formation that involves reduplication and voice prefixing, possibly with homorganic nasal assimilation. The sublexical rules are defined to capture word-internal hierarchical structures. The specification in the morphotactic part of the morphological analyser enables the morphological analyser to assign the tag +I in its output analysis. The tag is then able to be identified in XLE with its relevant features via the lexical entry for the suffix shown in (3a) for further processing, in particular for the identification of *–i* as an applicative or causative suffix. This is determined by classes of the stems and the annotation of the sub-lexical rules with relevant templates, indicated by @) in (4).

(4) V_STEM' --> { V_STEM-CAUSkan : @(VOICE @(CAUSkan VCausKan)); V_Kan_BASE
| V_STEM-CAUS_I : @(VOICE @(CAUS_I VCaus_I)) ; V_I_BASE
| V_STEM-APPLkan : @(VOICE @(APPLkan VAppKan)) ; V_Kan_BASE
| V_STEM-APPL_I : @(VOICE @(APPL_I VApp_I)) ; V_I_BASE
}.

```
(5)  VOICE(_SCHEMATA) = { _SCHEMATA @ACTOR-VOICE
                        | _SCHEMATA @UNDERGOER-VOICE
                                (^ OBJ) -->(^ SUBJ)
                                (^ SUBJ) -->(^ OBJ)
                        | _SCHEMATA @PASSIVE-VOICE
                                (^ OBJ) --> (^ SUBJ)
                                { SUBJ) --> NULL
                                | (^SUBJ) --> (^ OBL)
                                | (^SUBJ) --> (^ OBJ)
                                }
                        }.
```

(6) a. *Saya men-duduk-i kursi*
1s AV-sit-APPL chair
'I sat on a chair'

(6) b.



(6) c.

```
"Saya menduduki kursi"

      [PRED    'VApp_I<[6:saya], [26:kursi], 'duduk<[6:saya]>'>'
       SUBJ    [PRED  'saya'
                GLOSS [TRANS pronominal]
              6 NUM sg, PERS 1, PRON-TYPEpers]
       OBJ     [PRED  'kursi'
                GLOSS [TRANS chair]
                NTYPE [NSEM [COMMON count]
                       NSYN common ]
             26 PERS  3]
       GLOSS   [TRANS sit]
       TNS-ASP [MOOD indicative]
      12 APPLICATIVE+, CLAUSE-TYPEdecl, VOICE-TYPEAV

       [PRED    'duduk<[6:saya]>'
       SUBJ     [6:saya]
       GLOSS    [12-GLOSS]
       TNS-ASP  [12-TNS-ASP]
      -3 APPLICATIVE+, CLAUSE-TYPEdecl, VOICE-TYPEAV
```

The rules and their annotations regulate predicate compositions, i.e. constraining and/or supplying the relevant syntactic and semantic/aspectual information. For example, the template annotation of @(VOICE @(APPL_I VApp_I)), with internal bracketing, determines how applicativisation interacts with voice selection: predicates are first composed by the applicativisation according to the specification shown in (2), and then Voice selection constraints are applied. The Voice constraints called for by the templates are associated with voice types in Indonesian shown in (5), namely the AV (Actor Voice), UV (Undergoer Voice), and PASS (Passive voice). Their entries are specified in (3b-d).

(6b-c) show the parse output of (6a). The sublexical *c-str* tree of the verb in (6b) shows that XLE correctly identifies *menduduki* as a verb (with the +Verb tag suffix). *Men-* is correctly identified as the AV prefix (carrying the AV+ tag), signalling AV as indicated by the presence of the [VOICE-TYPE AV] feature in the *f-str* shown in (6c). The sublexical structure is not flat. This is linguistically motivated by the word formation system in Indonesian, and is also practically useful, e.g. the rule for V_STEM' (where the suffix *-i* (with the tag +I) is part of) needs to be specified only once for all types of voice selections. The *f-str* output in (6c) shows the analysis of the applicative suffix as a three-place predicate with the matrix SUBJ (tag no 6) shared with the SUBJ of the embedded predicate and *kursi* (tag no 26) a newly introduced OBJ argument. That the verb imposes an applicative structure is indicated by the presence of the [APPLICATIVE +] feature in the *f-str*.

Although there have been other works that have implemented valence changing strategies via predicate composition strategies using the XLE suite, our work on Indonesian is the first implementation of predicate composition for applicativisation, which takes into account language-specific voice selection.

REFERENCES

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar*. Standford: CSLI Publications.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Standford: CSLI Publications.

Butt, Miriam and King, Tracy Holloway. 2006. Restriction for morphological valency alternations: the Urdu causative. In Kaplan (ed.) *Intelligent Linguistic Architectures: Variations on Themes*. Stanford: CSLI Publications.

Butt, Miriam and King, Tracy Holloway King, John T. Maxwell III. 2003. Complex predicates via restrictions. In Butt and King (eds), *The proceedings of the LFG'03 Conference*. Stanford: CSLI Publications. http://csli-publications.stanford.edu/LFG/8/lfg03.html.

Chung, Sandra. 1976. On the subject of two passives in Indonesian. In: Charles N. Li (ed.), *Subject and topic* 57-99. NewYork: Academic Press.

Macdonald, R. Ross and Dardjowidjojo, Soenjono. 2001. *A student's reference grammar of modern formal Indonesian*. Georgetown: Georgetown University Press.

Musgrave, Simon. 2001. Non-subject arguments in Indonesian, PhD thesis, Melbourne University.

Myhill, J. 1988. Agent incorporation in Indonesian. *Journal of Linguistics* 24:111-136.

Purwo, Bambang Kaswanti. 1989. Voice in Indonesian : A Discourse Study. In B.K. Purwo (ed.), *Serpih -serpih telaah pasif bahasa Indonesia* 344-442. Jogyakarta: Kanisius.

Sneddon, James N. 1996. *Indonesian reference grammar*. St Leonards: Allen and Unwin.