Reduplication and Grammar Engineering for Indonesian

I Wayan Arka
ANU/Udayana University
wayan.arka@anu.edu.au

Ruli Manurung
Universitas Indonesia
maruli@cs.ui.ac.id

Meladel Mistica
ANU
meladel@gmail.com

Abstract

Reduplication is known to have posed a challenge in linguistic analysis and grammar engineering. Linguistically, issues include precise identification of its form-meaning relation. Formally, reduplication in Indonesian can interact with different kinds of affixation, which may involve morphophonemic alternations, e.g. segment deletion and/or nasal assimilation. This results in full and partial reduplication (1)-(2). Partial reduplication may also involve a change in the vowel segment, e.g. *lekak-lekuk* in (2).

(1) *pukul* (V) → *pukul-pukul* (V), *pukul-memukul* (V), *memukul-mukul* (V), *dipukul-pukul* (V), *berpukul-pukulan* (V), *dipukul-pukulkan* (V), *pukulan-pukulan* (N).
(2) *lekuk* (N) → *lekuk-lekuk* (N), *lekak-lekuk* (N), *berlekak-lekuk* (V)

Reduplication can express a range of meanings, and for simplicity, we will address its most common or core meaning in this paper, namely its 'plural' (or 'mass') meaning assignment. While we agree that Indonesian does not have a grammatical category of number as found in Indo-European languages such as English, we argue that at the functional/semantic level a linguistic form/structure carrying a plural meaning should be represented as having a feature structure of [NUM plural]. We will discuss simple cases of plural meaning assignment in reduplicated nouns, e.g. as in (3)a, but also cases of 'transfer' of plural meanings via predicate reduplication as in (3)b.

(3) a. *Gedung-gedung=nya besar* 'the buildings are huge'
    b. *Gedung=nya besar-besar* 'the buildings are huge'

We will report the progress we made in grammar engineering for Indonesian in implementing the analysis of reduplication outlined above within the LFG-based XLE platform. We will demonstrate the general aspects of how our Indonesian tokenizer and morphological analyser work as part of a broader system of the electronic grammar system. We will also discuss the challenge in handling different kinds of reduplication in a FST (finite-state transducer)-based morphological analyser, particularly how reduplication interacts with other affixation processes, and how such a non-concatenative operation is implemented on finite-state machinery using the "compile-replace" algorithm. We will address specific issues of architectural designs in grammar engineering that affect accuracy and efficiency, particularly in parsing sentences with reduplicated forms.