Productivity and Paradigmatic Relations:
A Corpus Based Analysis of Affix Substitution in Indonesian

Keywords: Indonesian prefixes, allomorphs, semantics roles, paradigmatic relations

Abstract
The morphology of Indonesian is characterized by productive processes of affix substitution. In this research, we are interested in two prefixes that create nouns from verbs and express a range of semantic functions (e.g. agent, instrument, patient; (Sneddon et al., 2010)). One prefix, henceforth pe- is assumed to attach to verbs with the prefix ber- by a process of affix substitution (e.g. petani ‘farmer’ – bertani ‘to farm’). A second prefix, henceforth peN- attaches to verbs with the prefix meN- (e.g. penari ‘dancer’ – menari ‘to dance’). This prefix has several allomorphs (peng-, peny-, pe-, pen-, pem-) that mirror the allomorphy of the base verbs (meng-, meny-, me-, men-, mem-, see (Ramlan, 2009) for details). However, corpus based investigation of these prefixes is still in its infancy, and no systematic surveys are available on the paradigmatic relations between the verbal base words and their corresponding derived nouns.

We compiled a database with detailed information on the pe- and peN- nouns and their base words, using the corpus of written Indonesian in the Leipzig Corpora Collection (Quasthoff et al., 2006). We analyzed the morphological structure of the words in this corpus using the MorphInd parser (Larasati et al., 2011) and checked results manually against the online version of Kamus Besar Bahasa Indonesia, a comprehensive dictionary of Indonesian (Alwi, 2012). Our database comprises a total of 3419 deverbal nouns (268 with pe-, 3146 with peN-, and 5 with the archaic variant per-), 5328 simple words (3503 nouns, 914 adjectives, and 910 verbs), as well as 12583 complex verbs (8915 with meN- and 3668 with ber-).

We probed the strength of paradigmatic relations between verbs and their corresponding derived nouns by investigating the correlation between their frequencies of occurrence in our corpus. For peN- and meN-, the Spearman correlation was 0.4054 (p < 0.0001), but for pe- and ber-, it was almost 0 (r_s = −0.0012, p = 0.9952). Figure 1 illustrates these correlations and also highlights that there are far fewer words with pe- than with peN-. Interestingly, our database contains 34 words where pe- is attached to verbs with prefixes other than ber- (12 words with di-, 12 words with meN-, 10 words with ter-). Thus, although pe- is most productive with verbs with ber-, its productivity is not constrained to exclusively verbs with ber-.

Furthermore, peN- appears primarily in words for agents and instruments, whereas pe- forms agent nouns and also shows some productivity for patient nouns (compare left and central panels in Figure 2). Finally, an analysis of the correlation between type counts and hapax legomena for the allomorphs of peN- shows that their productivity is proportional to the number of base verbs available. Pe-, however, does not participate in this proportionality (see right panel of Figure 2).

We conclude that word formation with pe- is less productive than word formation with peN-, and that words with pe- are more inviduated from their base forms compared to words with peN-. Our hypothesis is that a paradigmatic balance between base words and derived words, as reported here for affix substitution in Indonesian, is maintained in a language only in case the word formation process is fully productive.
Figure 1: Frequency of the noun as a function of the frequency of its base verb. The frequency of peN-nouns mirrors the frequency of the corresponding base words (left panel). No such correlation is present for pe- (right panel).

Figure 2: Semantics role counts of types and hapaxes for noun derived words. Pe-, written in 'Pe', is productive in forming patient and peN- allomorphs prefer instrument as their semantics specialization (compare left and central panels). Counts of types for base verbs (horizontal axis) and types and hapaxes for action nouns (vertical axis). The words with pe- are more inviduated from their base forms compared to allomorphs of peN- (right panel).

References


