How to deal with affixes in Wordnet Bahasa?
- a preliminary study -

David Moeljadi
D001@ntu.edu.sg

Division of Linguistics and Multilingual Studies,
Nanyang Technological University
Singapore

15 January 2016
Word formation in Indonesian

- Affixes: prefixes, infixes, suffixes, circumfixes
  - Inflectional: meN-, di-
  - Derivational
    - Noun-forming: peN-, pe-, per-, ter-, ke-, -el-, -em-, -er-, -an, -wan, -wati, -man, ke-...-an, peN-...-an, pe-...-an, per-...-an, ...
    - Adjective-forming: peN-, se-, ber-, ter-, -an, -i, -wi, -iah, meN-...-an, ...
    - Verb-forming: meN-, ber-, ter-, -kan, -i, ber-...-an, ber-...-kan, ter-...-kan, ke-...-an, ...
    - Adverb-forming: ter-, se-...-nya, ber-...-an, ...

- Clitics: ku=, =ku, kau=, =mu, =nya

- Reduplication: complete reduplication, partial reduplication, imitative reduplication
  - Reduplication + Affixes + Clitics

- Compounding: coordinative, subordinative
  - Compounding + Reduplication + Affixes + Clitics
Why should we care about this?

- Sense annotation in NTU-MC
  (http://compling.hss.ntu.edu.sg/ntumc/)

- Wordnet as a lexical source for DELPH-IN grammars, including Indonesian Resource Grammar (INDRA)
  (http://moin.delph-in.net/IndraTop)

- ...
Affixes in Wordnet Bahasa

Five groups:

1. Affixes *meN*- and *di-*
2. Clitics
3. Affixes not to be added to Wordnet
   - not to be separated from the stems
   - to be separated from the stems using lemmatizer
4. Affixes to be added to Wordnet (and separated from the stems)
5. Other affixes to be discussed

Evaluation/judgment criteria:

1. Productivity of the affixes
2. Consistency and predictability of meanings
3. Necessity of adding subtle meanings to Wordnet
Affixes *meN*- and *di*-

- **meN**: to be added to Wordnet together with the stems
  - Intransitive
    - the stems should not be added to Wordnet
  - Transitive
    - the stems should be added to Wordnet
    - a relation should be made between *stem* and *meN-*+stem
      (*stem* and *meN-*+stem refer to the same thing)

- **di**: not to be added to Wordnet and should be separated from the stems

→ We need information about transitive and intransitive verbs
Affix *meN-* in intransitive sentences

(1) a. *Adi tidur.*
    Adi sleep
    “Adi sleeps.”

b. *Tidur(-lah)!*
    “Sleep!”

(2) a. *Adi menangis.* (*meN-*+tangis)
    Adi cry
    “Adi cries.”

b. *Menangis(-lah)!*
    “Cry!”
Affixes *meN-* and *di-* in transitive sentences

(3) a. *Adi mengendar Budi.* (*meN-* + *kejar*)
   *Adi ACT-chase Budi*
   “Adi chases Budi.”

b. *Budi dikejar Adi.*
   *Budi PASS-chase Adi*
   “Budi is chased by Adi.”

c. *Budi saya kejar.*
   *Budi 1SG chase*
   “Budi is chased by me.”

d. *Kejar(-lah) Budi!*
   “Chase Budi!”
(4) a. \textit{Adi mem\textsubscript{perbaiki} sepeda. (meN-+per+-baik+-i)}
Adi \textsc{act}-repair bicycle

“Adi repairs a bicycle.”

b. \textit{Sepeda itu di\textsubscript{perbaiki} Adi.}
bicycle that \textsc{pass}-repair Adi

“That bicycle is repaired by Adi.”

c. \textit{Sepeda ini saya \textsubscript{perbaiki}.}
bicycle this \textsc{1sg} repair

“This bicycle is repaired by me.”

d. \textit{Perbaiki(-lah) sepeda itu!}

“Repair that bicycle!”
Affix *meN- + “instrument” noun stem*

(5) a. *Gunting ini tajam.*
    scissors this sharp
    “This pair of scissors is sharp.”

b. *Gunting(-lah) kertas ini!*
    scissor(-LAH) paper this
    “Cut this piece of paper with scissors!”

c. *Adi menggunting kertas. (meN-+gunting)*
    Adi ACT-scissor paper
    “Adi cuts a piece of paper with scissors.”

d. *Menggunting kertas adalah hobinya.*
    Scissoring paper COP hobby=3SG
    “Cutting paper with scissors is his/her hobby.”
Clitic ku= and =ku

- First person pronominal clitic “aku” 80001132-n

(6) a. *Budi kukejar.*
    Budi 1SG=chase
    “Budi is chased by me.”

b. *Budi telah memanggilku.*
    Budi PERF ACT-call=1SG
    “Budi has called me.”

c. *Buku ku jatuh.*
    book=1SG fall
    “My book falls.”
Clitic kau= and =mu

- Second person pronominal clitic “engkau”, “kamu” 77000023-n

(7) a. Budi kau kejar.
   Budi 2sg=chase
   “Budi is chased by you.”

b. Budi telah memanggil mu.
   Budi PERF ACT-call=2sg
   “Budi has called you.”

c. Buku mu jatuh.
   book=2sg fall
   “Your book falls.”
Third person pronominal enclitic “dia” 80001130-n

(8) a. *Budi dikejar*nya.
   Budi pass-chase=3SG
   “Budi is chased by him/her.”

b. *Budi telah memotong*nya.
   Budi perf act-cut=3SG
   “Budi has cut it.”

c. *Buku*nya jatuh.
   book=3SG fall
   “*His/her* book falls.”
Clitic =nya (2/3)

- Definiteness marker “the”

(9) *Buku* =nya jatuh.
    book=DEF fall
    “The book falls.”

- In a topic-comment clause

(10) *Gajah hidung* =nya panjang.
    elephant nose=NYA long
    “As for elephant, its nose is long.”
Clitic =nya (3/3)

- Nominalizer

(11) *Terbakar*nya rumah itu membuat warga panik.

burning=NYA house that ACT-make people panic

“The burning of that house makes people panic.”

- Adverbial marker: biasanya (normally), akhirnya (finally), sayangnya (pitiably), …
Affixes not to be added to Wordnet (1/3)

1. peN- : penulis (writer), penggaris (ruler), pencetak (printer), ...
2. pe- : pekerja (worker), peserta (participant), petinju (boxer), ...
3. per- : pelajar (learner), perantara (mediator), pertapa (hermit), ...
4. -an : makanan (food), balasan (reply), lapangan (field), lautan (ocean), manisan (sweetmeat), bulanan (monthly), mingguan (weekly), ...
5. ke-...-an : kedatangan (arrival), kedudukan (position), kesultanan (sultanate), kepulauan (archipelago), kepemimpinan (leadership), keteranganan (description), keharusan (must), kesatuan (unity), ...
6. peN-...-an : pemandanganan (scenery), penginapan (lodging), penggorengan (frying pan), ...
7. per-...-an : pergerakan (movement), pertemuuan (meeting), permusuhan (hostility), perpustakaan (library), persoalan (problem), pertigaan (T-junction), ...
1 pe-...-an : pedalaman (inland), pegunungan (highlands), petualangan (adventure), ...
2 ter- : terdakwa (defendant), tersangka (the suspect), terhukum (the sentenced), ...
3 pra- : prasejarah (prehistory), prasangka (prejudice), prakata (foreword), ...
4 -el- : telapak (palm), telunjuk (index finger), ...
5 -em- : gemetar (tremble), ...
6 -er- : seruling (flute), gerigi (serration), ...
7 meN-...-kan : menguntungkan (profitable), membosankan (boring), ...
Affixes not to be added to Wordnet (3/3)

1. -i : alami (natural), ...
2. -wi : duniawi (worldly), ...
3. -iah : rohaniah (spiritual), ...
4. tuna- : tuna rungu (deaf), tuna wisma (homeless), ...
5. ...

Moeljadi (LMS, NTU)
Affixes not to be added to Wordnet but to be separated using lemmatizer (1/2)

- **-an**: ribuan (thousands), 1920-an (1920s), ...
- **ke-**: keempat (all four), kelima (all five), ...

(12) *Keempat anak itu sangat nakal.*
KE-four kid that very naughty

“All those four kids are very naughty.”
Affixes not to be added to Wordnet but to be separated using lemmatizer (2/2)

- ber- : bertiga (all three), berenam (six together), ...

(13) _Mereka berenam teman saya_.
3PL BER-six friend 1SG

“They _all_ six (six _together_) are my friends.”

- ber-...-kan : beranggotakan (have ...as members), beristrikan (have ...as wife), ...

(14) _Dia beristri(kan) orang Minang_.
3SG BER-wife(-KAN) person Minang

“He _has_ a Minang person _as_ his wife.”
Affixes to be added to Wordnet (1/2)

- **se-**:
  - “same” 02068476-a: sekelas (same class), sekantor (same office), sependapat (same opinion), searah (same direction), ...
  - “whole” 00514884-a: se-Asia (whole Asia), se-Indonesia (whole Indonesia), sedunia (whole world), sekampung (whole village), ...
  - “as ...as” 00022131-r: secantik (as beautiful as), ...
  - “one” 13742573-n: seperlima (one fifth), ...
  - “one” 02186338-a: seekor (one CL), sekawanan (a herd), sekilo (one kilo), ...

- **-an** “more” 0009934-r: bagusan (better), gedean (bigger), cantikan (more beautiful), ...

- **ter-** “most” 00111609-r: terbaik (best), terjelek (ugliest), termahal (most expensive), ...
ke-…-an:

- “too” 00047392-r: kemahalan (too expensive), kekecilan (too short), ketuaan (too old), ...
- “rather”, “-ish” 00018781-r: kebiruan (bluish), kemerahan (reddish), kehijauan (greenish), ...

ke- linked to “ordinal” 02200035-a: kesatu (first), ketiga (third), ke-99 (99th), ...

antar- “antara” 80001161-n: antarbangsa (between nations), antarpulau (inter-island), antarbudaya (intercultural), ...
Other affixes to be discussed (1/3)

**dis-/un-...-ness**

- **ke-...-an** with *tidak* "NEG" : *ketidakmengertian* (*incomprehension*), *ketidakterbukaan* (*lack of openness*), *ketidakberdayaan* (*powerlessness*), *ketidakjujuran* (*dishonesty*), *ketidakberhasilan* (*unsuccessfulness*), ...
  - not listed in Kamus Besar Bahasa Indonesia (KBBI)

**-ing**

- **peN-...-an** : *pengiriman* (*sending*), *penerjemahan* (*translating*), *penandatanganan* (*signing*), *pengujian* (*testing*), ...
Other affixes to be discussed (2/3)

**Gender and number**

- **-wan**: wartawan (journalist), wisudawan (graduate), ...
- **-wati**: wartawati (female journalist), wisudawati (female graduate), seniwati (female artist), ...
- **-at**: muslimat (Muslim woman), mukminat (female believer), hadirat (female audience), ...
- **-in**: muslimin (Muslims), mukminin (believers), hadirin (audience), ...
Other affixes to be discussed (3/3)

Verb frames

- **-i** : *kirimi* (send 01031256-v), …
  → Somebody —s somebody something

(15) *Adi mengirim* _i saya surat._
    Adi ACT-send-I 1SG letter

  “Adi sent me a letter.”

- **-kan** : *kirimkan* (send 01031256-v), …
  → Somebody —s something
  → Somebody —s something to somebody

(16) *Adi mengirim* _kan surat (untuk saya)._*
    Adi ACT-send-KAN letter (to 1SG)

  “Adi sent a letter (to me).”
Other affixes to be discussed: ber-

- “have” 02203362-v: berbakat (talented), beruntung (lucky), beruang (moneyed), beristri (married, lit. have wife), ...
- “wear” 00050652-v: berdasi (wear a necktie), bertopi (wear a hat), berpayung (with an umbrella), berbedak (powdered), ...
- “ride”: bersepeda (ride a bicycle), berkuda (ride a horse), berperahu (ride a boat), ...
- “produce” 01752495-v: berbunga (flower, lit. produce flowers), bersuara (produce a voice), berbau (smell, lit. produce smell), ...
- To work as/in/at: berkuli (work as a coolie), berladang (work in an unirrigated field), ...
- Engange in an activity: berjemur (sunbathe), berkelahi (fight), beregang (swim), berpiknik (picnic), ...
Other affixes to be discussed: ber-

- “each other” : bertetangga (be neighbors), bersaudara (be brothers), berteman (be friends), ...
- to call someone with the name indicated by the noun or use it as a term of address : beraku (use aku “I”), berpaman (call someone as paman “uncle”), ...
- ...

Moeljadi (LMS, NTU)
Do we need to represent subtle differences? (1/2)

(17) *Buku itu ditinggalkan di perpustakaan.*
book that PASS-leave in library

“That book was left *(deliberately)* in the library” (Alwi et al. 2014: 151)

(18) *Buku itu ketinggalan di perpustakaan.*
book that KE-leave-AN in library

“That book was left *(accidentally)* in the library” (Alwi et al. 2014: 151)
Do we need to represent subtle differences? (2/2)

(19) *Perampok itu ditembak berkali-kali tetapi tidak tertembak.*
robber that PASS-shoot repeatedly but NEG TER-shoot

“That robber was shot at repeatedly but (he) was not shot” (Arka 2003: 10)

(20) *Rumah itu tidak dibakar tetapi terbakar.*
house that NEG PASS-burn but TER-burn

“That house was not burnt deliberately but (it) got/was burnt accidentally”