

# 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 **men**angis.* (***meN-**+tangis*)  
Adi cry  
“Adi cries.”
- b. ***Men**angis(-lah)!*  
“Cry!”

## Affixes *meN-* and *di-* in transitive sentences

(3) a. *Adi mengejar 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!”

## Affix *meN-* + *per-*

- (4) a. *Adi **mem**perbaiki sepeda.* (*meN-*+*per-*+*baik*+*-i*)  
Adi ACT-repair bicycle  
“Adi repairs a bicycle.”
- b. *Sepeda itu **di**perbaiki Adi.*  
bicycle that PASS-repair Adi  
“That bicycle is repaired by Adi.”
- c. *Sepeda ini saya perbaiki.*  
bicycle this 1SG repair  
“This bicycle is repaired by me.”
- d. *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. *Bukuku 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.”

## Clitic =nya (1/3)

- Third person pronominal enclitic “dia” 80001130-n

(8) a. *Budi dikejanya.*

Budi PASS-chase=3SG

“Budi is chased **by him/her.**”

b. *Budi telah memotongnya.*

Budi PERF ACT-cut=3SG

“Budi has cut **it.**”

c. *Bukunya jatuh.*

book=3SG fall

“**His/her** book falls.”

## Clitic =nya (2/3)

- Definiteness marker “the”

(9) *Bukunya jatuh.*

book=DEF fall

“**The** book falls.”

- In a topic-comment clause

(10) *Gajah hidungnya panjang.*

elephant nose=NYA long

“As for elephant, **its** nose is long.”

## Clitic =nya (3/3)

- Nominalizer

- (11) *Terbakarnya 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-** : **pen**ulis (writer), **peng**garis (ruler), **pen**cetak (printer), ...
- 2 **pe-** : **pe**kerja (worker), **pe**serta (participant), **pe**tinju (boxer), ...
- 3 **per-** : **pe**lajar (learner), **per**antara (mediator), **per**tapa (hermit), ...
- 4 **-an** : **ma**kanan (food), **ba**lasan (reply), **la**ngan (field), **la**utan (ocean), **ma**nisan (sweetmeat), **bu**lanan (monthly), **mi**ngguan (weekly), ...
- 5 **ke-...-an** : **ke**datangan (arrival), **ke**dudukan (position), **ke**sultanan (sultanate), **ke**pulauan (archipelago), **ke**pemimpinan (leadership), **ke**terangan (description), **ke**harusan (must), **ke**satuan (unity), ...
- 6 **peN-...-an** : **pe**mandangan (scenery), **pe**nginapan (lodging), **pe**nggorengan (frying pan), ...
- 7 **per-...-an** : **per**gerakan (movement), **per**temuan (meeting), **per**musuhan (hostility), **per**pustakaan (library), **per**soalan (problem), **per**tigaan (T-junction), ...

## Affixes not to be added to Wordnet (2/3)

- 1 **pe-...-an** : **pe**dalam**an** (inland), **pe**gunung**an** (highlands), **pe**tualang**an** (adventure), ...
- 2 **ter-** : **ter**dakwa (defendant), **ter**sangka (the suspect), **ter**hukum (the sentenced), ...
- 3 **pra-** : **pra**sejarah (**pre**history), **pra**sangka (**pre**judice), **pra**kata (**fore**word), ...
- 4 **-el-** : **te**lapak (palm), **te**lunjuk (index finger), ...
- 5 **-em-** : **ge**metar (tremble), ...
- 6 **-er-** : **se**ruling (flute), **ge**rigi (serration), ...
- 7 **meN-...-kan** : **meng**untung**kan** (profit**able**), **mem**bos**kan** (boring**g**), ...



## Affixes not to be added to Wordnet (3/3)

- ① **-i** : alami**i** (natural), ...
- ② **-wi** : dunia**wi** (worldly), ...
- ③ **-iah** : rohan**iah** (spiritual), ...
- ④ **tuna-** : tunarungu (deaf), tunawisma (homeless), ...
- ⑤ ...

## 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-** : **ber**tiga (**all** three), **ber**enam (six **together**), ...

(13) Mereka **ber**enam teman saya.

3PL BER-six friend 1SG

“They **all** six (six **together**) are my friends.”

- **ber-...-kan** : **ber**anggotakan (**have** ...**as** members), **ber**istrikan (**have** ...**as** wife), ...

(14) Dia **ber**istri(**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), se Kantor (**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), ...

## Affixes to be added to Wordnet (2/2)

- **ke-...-an:**
  - ▶ “too” 00047392-r : **ke**mahal**an** (**too** expensive), **ke**kecil**an** (**too** short), **ke**tua**an** (**too** old), ...
  - ▶ “rather”, “-ish” 00018781-r : **ke**biru**an** (blu**ish**), **ke**merah**an** (redd**ish**), **ke**hijau**an** (green**ish**), ...
- **ke-** linked to “ordinal” 02200035-a : **ke**satu (f**irst**), **ke**tiga (th**ird**), **ke**-99 (99**th**), ...
- **antar-** “antara” 80001161-n : **antar**bangsa (**between** nations), **antar**pulau (**inter**-island), **antar**budaya (**inter**cultural), ...

# Other affixes to be discussed (1/3)

## dis-/un-...-ness

- **ke-...-an** with *tidak* "NEG" : **ketidak**mengertian (*incomprehension*), **ketidak**terbukaan (*lack of openness*), **ketidak**berdayaan (*powerlessness*), **ketidak**jujukan (*dishonesty*), **ketidak**berhasilan (*unsuccessfulness*), ...  
→ not listed in Kamus Besar Bahasa Indonesia (KBBI)

## -ing

- **peN-...-an** : **pe**ngirim**an** (*sending*), **pe**nerjemah**an** (*translating*), **pe**n**andatangan****an** (*signing*), **pe**nguj**ian** (*testing*), ...

## Other affixes to be discussed (2/3)

### Gender and number

- **-wan** : warta**wan** (journalist), wisuda**wan** (graduate), ...
- **-wati** : warta**wati** (**female** journalist), wisuda**wati** (**female** graduate), seni**wati** (**female** artist), ...
- **-at** : muslim**at** (Muslim **woman**), mukmin**at** (**female** believer), hadir**at** (**female** audience), ...
- **-in** : muslim**in** (Muslim**s**), mukmin**in** (believer**s**), hadir**in** (audience), ...

## Other affixes to be discussed (3/3)

### Verb frames

- **-i** : kirim**i** (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** : kirim**kan** (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 : **ber**bakat (talented), **ber**untung (lucky), **ber**uang (moneyed), **ber**istri (married, lit. **have** wife), ...
- “wear” 00050652-v : **ber**dasi (**wear** a necktie), **ber**topi (**wear** a hat), **ber**payung (**with** an umbrella), **ber**bedak (powdered), ...
- “ride” : **ber**sepeda (**ride** a bicycle), **ber**kuda (**ride** a horse), **ber**perahu (**ride** a boat), ...
- “produce” 01752495-v : **ber**bunga (flower, lit. **produce** flowers), **ber**suara (**produce** a voice), **ber**bau (smell, lit. **produce** smell), ...
- to work as/in/at : **ber**kuli (work as a coolie), **ber**ladang (work in an unirrigated field), ...
- engage in an activity : **ber**jemur (sunbathe), **ber**kelahi (fight), **ber**enang (swim), **ber**piknik (picnic), ...

## Other affixes to be discussed: ber-

- “each other” : **ber**tetangga (be neighbors), **ber**saudara (be brothers), **ber**teman (be friends), ...
- to call someone with the name indicated by the noun or use it as a term of address : **ber**aku (use *aku* “I”), **ber**paman (call someone as *paman* “uncle”), ...
- ...

## 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**”