

Panduan Analisis Kata untuk Kamus Morfologi *MALINDO Morph*

Hiroki Nomoto, Livie Olivia, Thathit Puspaning Gegana,
Lukman Hakim, Friska Gania Putri, Shabrina Hazimi Putri

Februari 2018

revisi 1: September 2018

Hak cipta dilindungi © 2018 Universitas Kajian Asing Tokyo (東京外国語大学)

Daftar Isi

1.	Pendahuluan	2
2.	Kriteria pengecekan	2
2.1.	Bentuk reduplikasi yang ditulis dengan angka “2”	2
2.2.	Kata dasar merupakan kata majemuk	2
2.3.	Analisis kata dasar yang tidak tepat	3
2.4.	Analisis afiks yang tidak tepat	3
2.5.	Preposisi (<i>di, ke, dsb.</i>) yang dieja dengan nomina	3
2.6.	Kesalahan ejaan pada kata yang muncul di korpus	4
2.7.	Kegagalan analisis akibat huruf yang berubah menjadi angka	4
2.8.	Bahasa alay	5
2.9.	Kata-kata yang tidak biasa	5
2.10.	Kata-kata yang tidak dapat diidentifikasi	5
2.11.	Bentuk terikat (<i>antar-, anti-, multi-, non-, dsb.</i>)	6
2.12.	Partikel <i>pun</i>	6
2.13.	Kata tertulis dalam ejaan tidak baku atau ejaan lama	6
2.14.	Kata berprefiks <i>N-</i>	7
2.15.	Kata-kata yang disingkat penulisannya	7
2.16.	Penentuan kata dasar pada KD dan KBBI yang berbeda	7
2.17.	Penentuan kata dasar untuk kata-kata yang tidak selesai diucapkan	8

1. Pendahuluan

Panduan ini menjelaskan tentang kriteria-kriteria yang menjadi dasar korektor untuk mengesahkan, memilih dan mengoreksi hasil analisis yang dilakukan secara otomatis dengan penganalisis morfologi *MALINDO Morph*. Yang termasuk tugas korektor adalah

- mengesahkan analisis yang betul;
- menghapus analisis yang salah; dan
- mengoreksi analisis yang hampir betul supaya menjadi betul.

Notasi:

^ TAB

Bahan yang dirujuk dalam tugas pengoreksian:

1. Analisis morfologi *Kamus Dewan (KD)*
Pusat Rujukan Persuratan Melayu (<http://prpm.dbp.gov.my/>)
2. Analisis morfologi *Kamus Besar Bahasa Indonesia (KBBI)*
KBBI Daring (<https://kbbi.kemdikbud.go.id/Cari/Index>)
3. Daftar afiks bahasa Melayu
Bahasa Melayu TUFSS 文法
(http://www.tufs.ac.jp/common/fs/ase/mal/tatabahasa_web/affix_list.html)
[Selain afiks yang terkandung dalam daftar ini, bahasa Indonesia juga mempunyai prefiks *N-* dan sufiks *-in*.]

2. Kriteria pengecekan

2.1. Bentuk reduplikasi yang ditulis dengan angka “2”

Prakoreksi

```
couple2^ ber'couple2^ ber-^ 0^ 0^0
```

↓ (Menghapus angka “2” pada kata dasar dan mengganti angka “0” pada kolom reduplikasi menjadi “R-penuh”.)

Pascakoreksi

```
couple^ ber'couple2^ ber-^ 0^ 0^R-penuh
```

2.2. Kata dasar merupakan kata majemuk

Apabila kata yang muncul di korpus adalah kata majemuk atau kata yang terdiri dari dua kata

Prakoreksi

```
anaktir^ meng'anaktiri^ meN-^-i^0^ 0
```

```
anaktiri^ meng'anaktiri^ meN-^0^ 0^ 0
```

↓ (Menghapus baris pertama dan mengoreksi kata dasar pada baris kedua dengan menambahkan spasi di antara kedua kata.)

Pascakoreksi

```
anak tiri^ meng'anaktiri^ meN-^0^ 0^ 0
```

2.3. Analisis kata dasar yang tidak tepat

Apabila sebagian dari kata dasar secara tidak tepat dianalisis sebagai afiks

Prakoreksi

cur[^] di'curi[^] di-[^] -i[^]0[^] 0

↓ (Koreksi pada akar kata dan sufiks.)

Pascakoreksi

curi[^] di'curi[^] di-[^] 0[^] 0[^] 0

Prakoreksi

a[^] seakanku[^] se-[^] -kan+-ku[^] 0[^]0

↓ (Setelah mencari contoh penggunaannya di internet, “ku” dalam kata di atas merupakan subjek untuk kata berikutnya.)

Pascakoreksi

akan[^] seakanku[^] se-[^] -ku[^] 0[^]0

2.4. Analisis afiks yang tidak tepat

Apabila sebagian dari afiks secara tidak tepat dianalisis sebagai kata dasar

Prakoreksi

penghun[^] ber'penghuni[^] ber-[^] -i[^]0[^] 0

penghuni[^] ber'penghuni[^] ber-[^] 0[^] 0[^] 0

↓ (Koreksi pada kata dasar dan prefiks. Apabila terdapat beberapa afiks, ditautkan dengan tanda “+” sesuai dengan urutan abjad.)

Pascakoreksi

huni[^] ber'penghuni[^] ber-+peN-[^] 0[^] 0[^] 0

2.5. Preposisi (di, ke, dsb.) yang dieja dengan nomina

Apabila preposisi seperti di, ke, dan sebagainya secara tidak tepat dieja kata nomina, maka pada kolom prefiks diisi dengan preposisi tanpa tanda penghubung “-”

Prakoreksi

a[^] kedia[^] di-+ke-[^] 0[^] 0[^] 0

↓ (Koreksi pada akar kata dan prefiks.)

Pascakoreksi

dia[^] kedia[^] ke[^] 0[^] 0[^] 0

2.6. Kesalahan ejaan pada kata yang muncul di korpus

Apabila ejaan pada kata yang muncul di korpus tidak tepat, kata tersebut akan tetap disajikan apa adanya sehingga koreksi hanya dilakukan pada kolom kata dasar dan kolom lainnya

Prakoreksi

imara[^] meng'imarakan[^] meN-[^] -kan[^] 0[^] 0
imarak[^] meng'imarakan[^] meN-[^] -an[^] 0[^] 0

↓ (Kemungkinan maksud dari bentuk kata di atas adalah *mengiramakan*, sehingga pada kolom kata dasar diubah menjadi “irama”.)

Pascakoreksi

irama[^] meng'imarakan[^] meN-[^] -kan[^] 0[^] 0

Prakoreksi

kureng[^] ke'kurengan[^] 0[^] 0[^] ke--an[^] 0
kureng[^] ke'kurengan[^] ke-[^] -an[^] 0[^] 0
kurengan[^] ke'kurengan[^] ke-[^] 0[^] 0[^] 0

↓ (Kemungkinan maksud dari bentuk kata di atas adalah *kekurangan*, sehingga pada kolom kata dasar diubah menjadi “kurang”.)

Pascakoreksi

kurang[^] ke'kurengan[^] 0[^] 0[^] ke--an[^] 0

2.7. Kegagalan analisis akibat huruf yang berubah menjadi angka

Apabila ditemukan kegagalan analisis akibat huruf “o/O” yang menjadi angka “0” atau pada huruf “l/L” yang menjadi angka “1”

Prakoreksi

0rb[^] PENGORBANAN[^] 0[^] -an[^] peN--an[^] 0
0rbanan[^] PENGORBANAN[^] peN-[^] 0[^] 0[^] 0

↓ (Menganalisis sebagai *pengorbanan* dan melakukan koreksi pada kata dasar dan sufiks.)

Pascakoreksi

korban[^] PENGORBANAN[^] 0[^] 0[^] peN--an[^] 0

Prakoreksi

lum[^] sebelum[^] ber--se-[^] 0[^] 0[^] 0

↓ (Menganalisis sebagai *sebelum* dan melakukan koreksi pada kata dasar dan afiks.)

Pascakoreksi

belum[^] sebelum[^] se-[^] 0[^] 0[^] 0

2.8. Bahasa alay

Koreksi juga ditujukan terhadap bahasa alay dengan membetulkan kata dasar ke dalam bentuk yang bukan merupakan bentuk alay

Prakoreksi

4was[^] men4waskan[^] meN-[^] -kan[^] 0[^] 0
4waskan[^] men4waskan[^] meN-[^] 0[^] 0[^] 0

↓ (Mengubah “4was” menjadi “awas”.)

Pascakoreksi

awas[^] men4waskan[^] meN-[^] -kan[^] 0[^] 0

2.9. Kata-kata yang tidak biasa

Melakukan pencarian di internet atau kamus apabila bentuk kata yang muncul di korpus tidak diketahui dengan pasti

Prakoreksi

eh[^] Se'ehnya[^] 0[^] 0[^] se--nya[^] 0
eh[^] Se'ehnya[^] se-[^] -nya[^] 0[^] 0
ehnya[^] Se'ehnya[^] se-[^] 0[^] 0[^] 0

↓ (Setelah melakukan pencarian di internet, diketahui bahwa “se'eh” merupakan sebuah kata dalam dialek Pahang, Malaysia.)

Pascakoreksi

se'eh[^] Se'ehnya[^] 0[^] -nya[^] 0[^] 0

2.10. Kata-kata yang tidak dapat diidentifikasi

Apabila terdapat kata-kata yang meski sudah dilakukan pencarian lebih lanjut tetap tidak diketahui maknanya, maka bentuk kata yang muncul di korpus diperlakukan sebagai kata dasar

Prakoreksi

3l[^] di3lan[^] di-[^] -an[^] 0[^]0
3lan[^] di3lan[^] di-[^] 0[^] 0[^]0

↓ (Menganalisis kata “di3lan” sebagai kata dasar dan mengisi seluruh kolom di bagian kanan dengan angka “0”.)

Pascakoreksi

di3lan[^] di3lan[^] 0[^] 0[^] 0[^]0

Prakoreksi

aaa^ pengaaaaan^ peN-^ -an^ 0^ 0
aaa^ pengaaaaan^ 0^ 0^ peN--an^ 0

↓ (Menganalisis kata “pengaaaaan” sebagai kata dasar dan mengisi seluruh kolom di bagian kanan dengan angka “0”.)

Pascakoreksi

pengaaaaan^ pengaaaaan^ 0^ 0^ 0^ 0

2.11. Bentuk terikat (*antar-*, *anti-*, *multi-*, *non-*, dsb.)

Bentuk terikat seperti *antar-*, *anti-*, *multi-*, *non-*, dsb. dianalisis hanya apabila muncul bersamaan dengan afiks atau klitik

Prakoreksi

antarpema^ antarpemain^ 0^ -in^ 0^ 0

↓ (Memisahkan bentuk terikat dari kata dasar dan memasukkannya ke kolom prefiks.)

Pascakoreksi

main^ antarpemain^ antar-pe-^ 0^ 0^ 0

2.12. Partikel *pun*

Prakoreksi

sutawijayapun^ sutawijayapun^ 0^0^ 0^ 0

↓ (Membubuhkan spasi di antara partikel *pun* dan kata di depannya pada kolom kata dasar dan mengisikan “pun” tanpa tanda penghubung “-“ pada kolom sufiks.)

Pascakoreksi

sutawijaya pun^ sutawijayapun^ 0^pun^ 0^ 0

Prakoreksi

pertempuranpun^ pertempuranpun^ 0^ 0^ 0^ 0
tempuranpun^ pertempuranpun^ per-^ 0^ 0^ 0

↓ (Memisahkan partikel *pun* dan afiks dari kata dasar.)

Pascakoreksi

tempur^ pertempuranpun^ 0^ pun^ per--an^ 0

2.13. Kata tertulis dalam ejaan tidak baku atau ejaan lama

Prakoreksi

azaz^ Azazi^ 0^ -i^0^ 0
azazi^ Azazi^ 0^ 0^ 0^ 0

↓ (Membetulkan kata dasar ke ejaan baku.)

Pascakoreksi

asasi^ Azazi^ 0^ 0^ 0^ 0

Prakoreksi

sadja[^] Sadj^{a^} 0[^] 0[^] 0[^] 0

↓ (Membetulkan kata dasar ke Ejaan Bahasa Indonesia (EBI).)

Pascakoreksi

saja[^] Sadj^{a^} 0[^] 0[^] 0[^] 0

2.14. Kata berprefiks N-

Prakoreksi

fikir[^] Mikir[^] N-[^]0[^] 0[^] 0

pikir[^] Mikir[^] N-[^]0[^] 0[^] 0

↓ (Menghapus baris “fikir” yang merupakan kata dasar dalam bahasa Melayu dan menyisakan baris “pikir” yang merupakan bahasa Indonesia. Hal ini dikarenakan penggunaan prefiks N- yang umumnya ditemukan dalam bahasa Indonesia.)

Pascakoreksi

pikir[^] Mikir[^] N-[^]0[^] 0[^] 0

2.15. Kata-kata yang disingkat penulisannya

Prakoreksi

diaorg[^] diaorg[^] 0[^] 0[^] 0[^] 0

↓ (Membetulkan kata-kata di kolom kata dasar dengan menuliskan bentuk bakunya.)

Pascakoreksi

dia orang[^] diaorg[^] 0[^] 0[^] 0[^] 0

Prakoreksi

byk2[^] Byk2[^] 0[^] 0[^] 0[^] 0

↓ (Membetulkan kata-kata di kolom kata dasar dengan menuliskan bentuk bakunya.)

Pascakoreksi

banyak[^] Byk2[^] 0[^] 0[^] 0[^] R-penuh

2.16. Penentuan kata dasar pada KD dan KBBI yang berbeda

Dalam situasi yang demikian, analisis dipilih melalui sudut pandang linguistik. Pada kasus kata dasar menurut KD adalah *peluang* dan KBBI adalah *luang*.

Prakoreksi

peluang[^] peluang-peluang[^]0[^] 0[^]0[^] R-penuh

↓

Pascakoreksi

luang[^] peluang-peluang[^]pe-[^] 0[^]0[^] R-penuh

Kasus serupa juga muncul pada kata lainnya seperti:

- **Sehingga**

Kata dasar menurut *KD* adalah *hingga*, sedangkan menurut *KBBI* adalah *sehingga*. Dalam kasus ini, tidak dilakukan koreksi terhadap analisis di bawah karena *hingga* dinilai sebagai kata dasar yang sudah tepat.

hingga [^] sehing-ga [^] se- [^] 0 [^] 0 [^] 0

- **Laki-laki**

Kata dasar menurut *KBBI* adalah *laki*, sedangkan menurut *KD* adalah *laki* atau *laki-laki*. Dalam hal ini, *laki* dinilai sebagai kata dasar yang tepat.

<u>Prakoreksi</u> laki2 [^] laki2 [^] 0 [^] 0 [^] 0 [^] 0 ↓ <u>Pascakoreksi</u> laki [^] laki2 [^] 0 [^] 0 [^] 0 [^] R-penuh
--

2.17. Penentuan kata dasar untuk kata-kata yang tidak selesai diucapkan

Kata dasar ditulis berdasarkan bentuk utuhnya.

kutip [^] mengu [^] meN- [^] 0 [^] 0 [^] 0

Kalimat asli dalam korpus:

Pekejaan Mat Jenin adalah, **mengu** /// memanjat pokok kelape dan **mengutip** kelape untuk dijual di pasa-pasa. [KL201712, *Korpus Variasi Bahasa Melayu: Standard Lisan*]