Morphological Segmentation

Annotation Guidelines

Appendix: Indonesian

Linguistic Data Consortium

This document contains examples of common linguistic patterns in Indonesian that are relevant to the Morphological Segmentation task. They represent classes of morphological patterns in the language that could be annotated in the task. Note that these examples do no necessarily indicate what the annotation should be.

In particular, the notational conventions used in these descriptions are at odds with the conventions used in the annotation. For instance, reduplication is shown with a hyphen, whereas in the annotation the elements will be separated from each other.

> This sketch: *te-tangga*
> Annotation: *te - tangga*

Also note that prefixes that have the convention *-N* to mark a variable nasal should be treated as if the basic form is *–n* for ease of annotation. In cases where the surface form is *–ng*, the orthographic *g* is an addition and so does not have to be explicitly marked.

> pencetak "printer" = pen cetak
> pengukur "one who measures = peng ukur
> pembuat "maker" = pe<m> buat

Below are some paradigms of Indonesian morphological patterns.

	Root		Reduplicated form
i.	tangga 'stairs, ladder'	>	te-tangga 'neighbor'
ii.	berapa 'how many'	>	be-berapa 'a few'
iii.	tapi 'but'	>	te-tapi 'but'

Table 4.2: Partial reduplication

1.	buat 'make'	>	pembuat 'maker, the one who makes'
2.	cukur 'shave'	>	pencukur 'one who shaves, beard and mustache barber'
3.	tutup 'close'	>	penutup 'lid, cover'
4.	cetak 'print'	>	pencetak 'printer'
5.	malu 'shy'	>	pemalu 'shy person'
6.	sebab 'cause'	>	penyebab 'cause'
7.	ukur 'size'	>	pengukur 'one who measures; thing with which one measures'
8.	lari 'run'	>	pelari 'runner'
9.	saing 'compete'	>	pesaing 'competitor'
10.	tunjuk 'point'	>	penunjuk 'instructions'
		>	penunjuk 'indicator'
11.	lupa 'forget'	>	pelupa 'forgetful'
12.	marah 'angry'	>	pemarah 'hotheaded'

Table 4.3: Forms with **peN-**

	Root		Inflected form
13a.	cepat 'fast'	>	percepat 'make fast, hasten'
13b.	antara 'size'	>	perantara 'go-between'

Table 4.4: Forms with **per-**

	Root		Inflected form
14.	lukis 'paint'	>	lukisan 'painting'
15.	camil 'snack'	>	camilan 'snack'
16.	curi 'steal'	>	curian 'stolen'
17.	sewa 'rent'	>	sewaan 'rented'
18.	bantu 'help'	>	bantuan 'aid, assistance'
19.	sewa 'rent'	>	sewaan 'rented'
20.	landas 'base'	>	landasan 'base'
21.	laut 'sea'	>	lautan 'ocean'

Table 4.5: Forms with **-an**

	Root		Inflected form
22.	baik 'good'	>	kebaikan 'goodness'
23.	sehat 'healthy'	>	kesehatan 'health'
24.	datang 'come'	>	kedatangan 'arrival'
25.	besar 'big'	>	kebesaran 'too big'
26.	lambat 'late'	>	kelambatan 'too late'
27.	hujan 'rain'	>	kuhajanan 'get rained on'
28.	gelap 'dark'	>	kegelapan 'get overtaken by dark/night'
29.	berhasil 'succeed'	>	keberhasilan 'success'
30.	tersedia 'available'	>	ketersediaan 'availability'
31.	adil 'fair'	>	ketidak-adilan 'injustice'
32.	duduk 'sit'	>	penduduk 'population, inhabitant'

	Root		Inflected form
33.	cetak 'print'	>	pencetakan 'printing'
34.	jual 'sell'	>	penjualan 'sale of s.t.'
35.	pandang 'view'	>	pemandangan 'view'

Table 4.7: Forms with **peN-...-an**

	Root		Inflected form
36.	lawan 'oppose'	>	perlawanan 'opposition'
37.	kerja 'work'	>	pekerjaan 'work'
38.	gerak 'move'	>	perkerakan 'movement'
39.	tolong 'help'	>	pertolongan 'assistance'
40.	judi 'gamble'	>	perjudian 'place for gambling'
41.	cetak 'print'	>	percetakan 'printing house'
42.	ekonomi 'economy'	>	perekonomian 'economic'
43.	hotel 'hotel'	>	perhotelan 'related to hotel industry'
44.	batas 'border'	>	perbatasan 'border'
45.	masalah 'issue'	>	permasalahan 'issue'

Table 4.8: Forms with **per-...-an**

This prefix is used to derive abstract nouns from various bases.

	Root		Inflected form
46.	hendak 'wish'	>	kehendak 'desire'
47.	tua 'old'	>	ketua 'chief, head'

Table 4.9: Forms with **ke-**

	Root		Inflected form
48.	warta 'news'	>	wartawan / wartawati 'journalist'
49.	olahraga 'sports'	>	olahragawan / olahragawati 'athlete'

Table 4.10: Forms with **-wan, -wati**

	Root		Inflected form
50.	ekor 'tail'	>	seekor 'one tail' (used as a classifier for animals)
51.	rumah 'house'	>	serumah 'in the same house'
52.	dunia 'world'	>	sedunia 'the whole world'

Table 4.11: Forms with **se-**

	Root		Inflected form
53.	nganga 'agape'	>	senganga-nganganya 'as agape as possible'
54.	lama 'long' (time)	>	selama-lamanya 'as long as possible'

Table 4.12: Forms with **se-…-nya**

affix	function
meN-	agent/actor focus verbal prefix
di-	patient focus verbal prefix
ber-	intransitive verbal prefix
-kan	applicative suffix marking: transitivity, causativity
-i	applicative suffix marking: location, transitivity

Table 4.16: Verbal affixes