

Morphological Segmentation

Annotation Guidelines

Appendix: Tagalog

Linguistic Data Consortium

© [2018] Trustees of the University of Pennsylvania

*****This document is unpublished and intended solely for the use of the individual or entity to whom it was delivered. Redistribution is strictly prohibited without the express authorization of the Linguistic Data Consortium.*****

This document contains examples of common linguistic patterns in Tagalog 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 not 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, infixes here are labeled with angled brackets (<>), while in the annotation scheme, these are used to mark substituted elements.

This sketch: *b<um>ili*
Annotation: *b um ili*

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.

pantulog “sleepwear” = pan tulog
panghugas “washing tool” = pang hugas
pambili “money for purchasing something” = pa<m> bili

Reduplication is segmented off from the rest of the stem:

magsasaka “farmer” = mag sa saka

In cases where a consonant in the stem is affected by another segment, the changed consonant should be marked in angled brackets.

mangingisda “fisherman” = mang i<ng> isda
manunulat “writer” = man u <n>ulat

In the last example, it would be plausible to segment this as *ma <n>u <n>ulat* to show the change of the base form *sulat*. However, for ease of annotation, we will keep the prefix together.

Below are some paradigms of Tagalog morphological patterns.

1. *mag*+RED-, *maN*+RED-: Agent nominalization, similar to English *-er*, referring to one who typically does some action. Prefix is often homophonous with agent voice morphology, but stress on the entire form is often assigned differently.

³My intuition is that the perfective forms are more marginal. At the very least, the forms provided by Schachter and Otnes do not neatly correspond to the forms I would use in my dialect.

- (21) a. *mag* + RED + *saka* '(to) farm' → *magsasaka* 'farmer'
b. *maN* + RED + *isda* '(to) fish' → *mangingisda* 'fisherman'⁴
c. *maN* + RED + *laro* '(to) play' → *manlalaro* 'player'
d. *maN* + RED + *sulat* '(to) write' → *manunulat* 'writer'
2. *taga* -: Another agent nominalization, although forms with this prefix tend to have a meaning of "person who has the responsibility of doing X".

- (22) a. *taga* + *linis* '(to) clean' → *tagalinis* 'cleaner'
b. *taga* + *sulat* '(to) write' → *tagasulat* 'one in charge of writing'
c. *taga* + *pag* + *salita* '(to) speak' → *tagapagsalita* 'speaker/spokesperson'

This prefix also creates demonyms:

- (23) a. *taga* + *Canada* → *taga-Canada* 'Canadian/person from Canada'
b. *taga* + *Amerika* 'America' → *taga-Amerika* 'American/person from America'
c. *taga* + *Maynila* 'Manila' → *taga-Maynila* 'Manilan/person from Manila'

3. *-in*: Theme nominalization, creates a form referring to the object of the root verb. Note that this suffix is homophonous with the patient voice suffix *-in*. However, the forms that result from the application of either suffix differ in terms of stress placement as well as the potential for vowel deletion (no vowel deletion occurs with the nominalizer).

- (24) a. *gawa* 'do/make' + *in* → *gawain* 'activity'
b. *gastos* 'spend' + *in* → *gastusin* 'expense'
c. *bili* 'buy' + *in* → *bilihin* 'goods, things to buy'
d. *hugas* 'wash' + *in* → *hugasin* 'things to wash'

4. **-an**: Location nominalization, creates a form referring to a location where the event described by the root verb can or does occur. Note that, similarly to *-in* above, this suffix is homophonous with the locative voice suffix *-an*, with differences in the stress placement and vowel deletion potential in the produced forms

- (25) a. (*pa* +) *gawa* 'do/make' + *an* → (*pa*)*gawaan* 'factory'
 b. *bili* 'buy' + *an* → *bilihan* 'store, market'
 c. *hugas* 'wash' + *an* → *hugasan* 'washing place'
 d. *tanim* '(to) plant' + *an* → *taniman* 'planting place, plantation'

Also creates nominalizations referring to events, but not in the same way as gerunds.

- (26) a. *suntok* 'punch' + *an* → *suntukan* 'fistfight, brawl'
 b. *sayaw* 'dance' + *an* → *sayawan* 'a dance' (i.e., a ball, a party)
 c. *inom* 'drink' + *an* → *inuman* 'drinking get-together' (i.e., drinks at a bar/pub)

- (27) a. *paN* + *tulog* 'sleep' → *pantulog* 'sleepwear' (or *pangtulog*)
 b. *paN* + *sulat* 'write' → *panulat* 'writing instrument'
 c. *paN* + *hugas* 'wash' → *panghugas* 'washing tool'
 d. *paN* + *bili* 'buy' → *pambili* 'money for purchasing sth.' (or *pangbili*)

Abstract Nouns These forms generally refer to abstract concepts like *happiness* or *peace*. They are formed with a circumfix *ka-...-an*. This circumfix is potentially analyzable as two simpler prefixes, but it is not immediately clear how to do this.

- (28) a. *ka* + *payapa* 'peaceful' + *an* → *kapayapaan* 'peace'
 b. *ka* + *babae* 'woman' + *an* → *kababaihan* 'womankind, womanhood'
 c. *ka* + *giting* 'heroic' + *an* → *kagitingan* 'honor, heroism'
 d. *ka* + *sulat* 'write' + *an* → *kasulatan* 'writings, scripture'

(33) "Concrete" *ma*- adjectives

- a. *ma* + *bundok* 'mountain' → *mabundok* 'mountainous'
 b. *ma* + *tao* 'person' → *matao* 'crowded, populous'
 c. *ma* + *pera* 'money' → *mapera* 'moneyed, rich'

18

(34) "Abstract" *ma*- adjectives

- a. *ma* + *laki* 'big, size' → *malaki* 'big'
 b. *ma* + *galing* 'skillful, skill' → *magaling* 'skillful'
 c. *ma* + *yaman* 'rich, wealth' → *mayaman* 'rich, wealthy'

- (37) a. *malaki* 'big' vs *malalaki* 'big (pl.)'
 b. *maganda* 'beautiful' vs *magaganda* 'beautiful (pl.)'
 c. *mabundok* 'mountainous' vs *mabubundok* 'mountainous (pl.)'
 d. *mapera* 'moneyed, rich' vs **mapepera*

Table 10: Some intensified adjective forms

Subclass	Positive	Napaka-	Ang (RED _{stem})	ADJ LK ADJ
Concrete Ma- <i>mabundok</i> 'mountainous'	<i>ma</i> +STEM	<i>napaka</i> + <i>ma</i> +STEM <i>napakamabundok</i>	— —	<i>mabundok na mabundok</i>
Abstract Ma- <i>malaki</i> 'big'	<i>ma</i> +STEM	<i>napaka</i> +STEM <i>napakalaki</i>	<i>ang</i> STEM (STEM) <i>ang laki (laki)</i>	<i>malaking malaki</i>
Result Unaffixed <i>gising</i> 'awake'	STEM	<i>napaka</i> +STEM <i>?napakagising</i>	— —	<i>gising na gising</i>
Simple Unaffixed <i>payat</i> 'thin'	STEM	<i>napaka</i> +STEM <i>napakapayat</i>	<i>ang</i> STEM (STEM) <i>ang payat (payat)</i>	<i>payat na payat</i>

- (42) a. *P<in>untah-an ng guro ang party.*
 <PFV>go-LV GEN teacher NOM party
 'The teacher went to the party.' Locative Voice
- b. *P<um>unta ang guro sa party.*
 <AV.PFV>go NOM teacher OBL party
 'The teacher went to the party.' Agent Voice
- (43) a. *I-b<in>ili ng guro ng mangga ang estudyante.*
 CV-<PFV>buy GEN teacher GEN mango NOM student
 'The teacher bought a mango for the student.' Conveyance Voice
- b. *B<um>ili ang guro ng mangga para sa estudyante.*
 <AV.PFV>buy NOM teacher GEN mango for OBL student
 'The teacher bought a mango for the student.' Agent Voice

TABLE 12. Sample root aspect paradigm

	Infinitive	Perfective	Imperfective	Contemplative	Stem
<um>	<um>STEM <i>k<um>ain</i>	<um>STEM <i>k<um>ain</i>	<um>RED+STEM <i>k<um>akain</i>	RED+STEM <i>kakain</i>	<i>kain</i> 'eat'
<i>m</i> -initial	<i>mag</i> +STEM <i>maglaro</i>	<i>nag</i> +STEM <i>naglaro</i>	<i>nag</i> +RED+STEM <i>naglalaro</i>	<i>mag</i> +RED+STEM <i>maglalaro</i>	<i>laro</i> 'play'
	<i>maN</i> +STEM <i>mangalaga</i>	<i>naN</i> +STEM <i>nangalaga</i>	<i>naN</i> +RED+STEM <i>nangangalaga</i>	<i>maN</i> +RED+STEM <i>mangangalaga</i>	<i>alaga</i> 'care for'
	<i>mamili</i>	<i>namili</i>	<i>namimili</i>	<i>mamimili</i>	<i>bili</i> 'buy'
	<i>manloko</i>	<i>nanloko</i>	<i>nanloloko</i>	<i>manloloko</i>	<i>loko</i> 'fool'
	<i>ma</i> +STEM <i>mainom</i>	<i>na</i> +STEM <i>nainom</i>	<i>na</i> +RED+STEM <i>natinom</i>	<i>ma</i> +RED+STEM <i>mainom</i>	<i>inom</i> 'drink'
	<i>maka</i> +STEM <i>makagawa</i>	<i>naka</i> +STEM <i>nakagawa</i>	<i>nakaka</i> +STEM <i>naka</i> +RED+STEM ¹ <i>nakakagawa</i> <i>nakagagawa</i>	<i>makaka</i> +STEM <i>maka</i> +RED+STEM ¹ <i>makakagawa</i> <i>makagagawa</i>	<i>gawa</i> 'do'
Other	STEM+ <i>in</i> <i>tapusin</i> <i>lutuin</i>	< <i>in</i> >STEM <i>ni</i> +STEM ² <i>tinapos</i> <i>niluto</i>	< <i>in</i> >RED+STEM <i>ni</i> +RED+STEM ² <i>tinatapos</i> <i>niluluto</i>	RED+STEM+ <i>in</i> <i>tatapusin</i> <i>lulutuin</i>	<i>tapos</i> 'finish' <i>luto</i> 'cook'
	STEM+ <i>an</i> <i>puntahan</i> <i>lutuan</i>	< <i>in</i> >STEM+ <i>an</i> <i>ni</i> +STEM+ <i>an</i> ² <i>pinuntahan</i> <i>nilutuan</i>	< <i>in</i> >RED+STEM+ <i>an</i> <i>ni</i> +RED+STEM+ <i>an</i> ² <i>pinupuntahan</i> <i>nilulutuan</i>	RED+STEM+ <i>an</i> <i>pupuntahan</i> <i>lulutuan</i>	<i>punta</i> 'go' <i>luto</i> 'cook'
	<i>i</i> +STEM <i>isulat</i> <i>ilagay</i> <i>ihanda</i>	<i>i</i> +< <i>in</i> >STEM <i>ini</i> +STEM ³ <i>isinulat</i> <i>inilagay</i> <i>inihanda</i>	<i>i</i> +< <i>in</i> >RED+STEM <i>ini</i> +RED+STEM ³ <i>isinusulat</i> <i>inilalagay</i> <i>inilahanda</i>	<i>i</i> +RED+STEM <i>isusulat</i> <i>ilalagay</i> <i>ihahanda</i>	<i>sulat</i> 'write' <i>lagay</i> 'put' <i>handa</i> 'ready'

1. *pa*:- Causative; attaches to the root to result in forms indicating that the action or event described is caused by some external (usually sentient) entity.

- (50) a. <*in*> + *pa* + *baba* 'down, (to) go down' → *pinababa* 'caused sth. to go down'
b. *nag* + *pa* + *gawa* 'do/make' → *nagpagawa* 'caused sb. to do/make sth.'
c. *mag* + *pa* + *dala* 'bring' → *magpadala* 'to send'
d. *i* + *pa* + *kuha* 'get/take' → *ipakuha* 'to cause sb. to get/take sth.'

This prefix may also appear without an accompanying voice morpheme to produce a polite request form. This form does not inflect further.

- (51) a. *pa* + *gawa* 'do/make' → *pagawa* 'please do/make'
b. *pa* + *kuha* 'get/take' → *pakuha* 'please get/take'

2. *ka*:- Reason/Cause; occurs in locative and conveyance voices mostly with adjectival roots to indicate the (potentially non-sentient) reason for the state described by the root.

- (52) a. *i* + <*in*> + *ka* + *galit* 'angry' → *ikinagalit* 'was angered by sth.'
b. *ka* + RED + *tuwa* 'amused' + *an* → *katutuwaan* 'will be amused by sth.'

3. *mag-/pag-*: Typically indicates transitivity alternations of some sort (compare (53) and (54)), although interacts semantically in non-trivial ways with other affixes.

- (53) a. *mag* + *baba* 'down' → *magbaba* 'to bring/let down'
 b. *mag* + *bukas* 'open' → *magbukas* 'to open sth.'
- (54) a. <*um*> + *baba* 'down' → *bumaba* 'to go down'
 b. <*um*> + *bukas* 'open' → *bumukas* 'to open'
- (55) a. *pag* + *kuha* 'take' + *an* → *pagkunan* 'to take from sth./sb.'
 b. *pag* + *aral* 'learn' + *an* → *pag-aralan* 'to study'
 c. *maka* + *pag* + *aral* 'learn' → *makapag-aral* 'to be able to study'

4. *maN-/paN-*: Often conveys a state of some entity being affected in some way, carries some negative connotation, or denotes some instrumentality. Homorganic nasal sometimes triggers nasal substitution (see Sec. 4.8).

- (56) a. *maN* + *anak* 'child' → *manganak* 'to give birth'
 b. *maN* + *isda* 'fish' → *mangisda* 'to fish'
 c. *maN* + *kagat* 'bite' → *mangagat* 'to bite'
 d. *maN* + *yari* 'create' → *mangyari* 'to happen'
 e. *i* + *paN* + *bigay* 'give' → *ipamigay* 'to give sth. out'
 f. *i* + *paN* + *punas* 'wipe' → *ipampunas* 'to use for wiping'
5. *ma-*: Prominently marks the "Ability and Involuntary Action" form (Schachter and Otnes 1972, Ch. 5.13); also marks certain subclasses of intransitive verbs.
- (57) a. *ma* + *basa* 'read' → *mabasa* 'to be able to read'
 b. *ma* + *bigay* 'give' → *mabigyan* 'to be able to give'
 c. *ma* + *kain* 'eat' → *makain* 'to be able to eat'
- (58) a. *ma* + *tuwa* 'amused' → *matuwa* 'to become amused'
 b. *ma* + *tulog* 'sleep' → *matulog* 'to sleep'
 c. *ma* + *lamig* 'cold' + *an* → *malamigan* 'to feel cold'
6. *maka-*: Marks the "Ability and Involuntary Action" form, predominantly for agent voice forms.
- (59) a. *maka* + *kita* 'see' → *makakita* 'to be able to see (sth.)'
 b. *maka* + *kain* 'eat' → *makakain* 'to be able to eat'
 c. *maka* + *pag* + *luto* 'cook' → *makapagluto* 'to be able to cook'

7. *maki-/paki-*: Marks “Social” verbs (Schachter and Otnes 1972, Ch. 5.14), denoting that an action was carried out collectively or in some shared manner. Sometimes results in idiosyncratic meanings.

- (60) a. *maki* + *tulog* ‘sleep’ → *makitulog* ‘to share the sleeping facilities of sb.’
- b. *maki* + *pag* + *usap* ‘converse’ → *makipag-usap* ‘to converse with sb.’
- c. *maki* + *usap* ‘converse’ → *makiusap* ‘to petition/bargain with sb.’
- d. *paki* + *usap* ‘converse’ + *in* → *pakiusapin* ‘to petition/bargain with sb.’
- e. *maki* + *alam* ‘know’ → *makialam* ‘to meddle/interfere in sth.’
- f. *paki* + *alam* ‘know’ + *an* → *pakialaman* ‘to meddle/interfere in sth.’

8. *magsi-*: Marks pluralized verbs (see Sec. 4.5.2; Schachter and Otnes 1972, Ch. 5.15).

- (61) a. *magsi* + *tawa* ‘laugh’ → *magsitawa* ‘to laugh (as a group)’
- b. *nagsi* + RED + *pag* + *aral* ‘study’ → *nagsisipag-aral* ‘are (all) studying’

9. *magka-*: Under verbal uses, means “to come to have”.

- (62) a. *magka* + *anak* ‘child’ → *magkaanak* ‘to come to have a child’
- b. *magka* + *sakit* ‘pain/illness’ → *magkasakit* ‘to become sick’

4.8.1 Vowel Alternations

As noted in Sec. 2.5, there is some phonological alternation between /o~u/. Generally speaking, /o/ becomes /u/ when not in the final syllable of a word, as in (94). However, some counterexamples exist, as in (95).

- (94) a. *bago* ‘new’ + *in* → *baguhin* ‘to change PV’
- b. *bakod* ‘fence’ + *an* → *bakuran* ‘lawn, yard’
- c. *ka* + *gwapo* ‘handsome’ + *an* → *kagwapuhan* ‘handsomeness’
- (95) *ka* + *bobo* ‘stupid’ + *an* → *kabobohan* ‘stupidity’ (**kabobuhan*, **kabubuhan*)

The alternation is much less robust for /e~i/, although there is at least one example, shown in (96). There appear to be not many native origin roots in Tagalog that have *e* in their final syllable. Potential examples that take suffixes are of Spanish origin and do not seem to undergo this process, as in (97). This likely reflects the fact that the alternation was more robust (i.e., the high and mid vowels are argued to have been non-contrastive) before widespread contact with English and Spanish.

- (96) *ka* + *babae* ‘woman’ + *an* → *kababaihan* ‘womankind, womanhood’
- (97) *karne* ‘meat’ + *an* → *karnehan* ‘meat place (usu. butcher shop, but possibly also meat restaurant, etc.)’ (**karnihan*)

4.8.2 D–R Alternation

In words of native origin, *d* and *r* alternate with each other, with *r* appearing generally in between vowels and *d* appearing elsewhere. This alternation may be somewhat weaker in Modern Tagalog when compared to earlier stages of the language, likely due to contact with and borrowing from Spanish and English.

The alternation is most robust within words as in (98), although there appear to be exceptions as in (99).

- (98) a. *ma* + *dami* ‘amount’ → *marami* ‘many’ (although *madami* is occasionally attested)
b. *ka* + *dunong* ‘knowledge, wisdom’ + *an* → *karunungan* ‘skill’
c. *ka* + RED₂ + *dapat* ‘should, right’ → *karapat-dapat* ‘deserving, justified’
d. *na* + RED + *dinig* ‘hear’ → *naririnig* ‘hears, is hearing’
- (99) a. *ma* + *dulas* ‘slip, slippery’ → *madulas* ‘slippery, to slip’
b. *ma* + *damot* ‘stingy, miserly’ → *madamot* ‘stingy, miserly’ (*maramot* is listed in Almarino (2010), but the former appears to be more common)
c. RED + *dampot* ‘pick up from the ground’ + *in* → *dadamputin* ‘will pick up from the ground’ (*daramputin* appears to be rarely attested)

4.8.3 H-insertion with Vowel-final Roots

When appearing with suffixal material, vowel-final roots in Tagalog surface with an *h* at the end of the root.

- (100) a. *ka* + *babae* ‘woman’ + *an* → *kababaihan* ‘womankind, womanhood’
b. *bili* ‘buy’ + *in* → *bilihin* ‘goods/items to purchase’
c. *kantá* ‘sing’ + *an* → *kantáhan* ‘singing event/party’