# Formalization of Sinhalese Morphology

2E−2

S. Herath‡, T. Ikeda, S. Ishizaki , Y. Anzai‡

‡Keio University, 3 - 14 - 1, Hiyoshi, Yokohama 223
Electrotechnical Laboratory, 1 - 1 - 4, Umezono, Tsukuba-shi 305

## Introduction

We are developing a Sinhalese analysis system. In this paper, we describe the formalization of morphology as its basic step.

(note: Sinhalese is the major language in Sri Lanka, spoken by 15 million. It has its own alphabet. We have developed a romanization mapping for machine processing[1].

ජපානය
JaPaNaYa (Japan)

## 1 Morphological structure

The basic Sinhalese parts of speech are noun, verb, particle, and suffix. The noun, verb and particle forms a grammartical unit with or without suffixes. There is separation between units, but not between the components.

root suf suf suf = noun unit
GaS + A + Ta + Th = GaSaTaTh (to the tree too)
OHu GaSaTaTh NaGi*. (He climbed to the tree too.)

As in the above example, when forming a unit, there can be some character changes in the components. We formalized them as linking rules in section 4.

Noun unit can have at most three suffixes. Verb unit can have only one suffix. Verb root takes several different forms at linking with suffixes. We have formalized those forms as verb root inflections. Noun has no such inflection. Noun and verb units have attributes such as gender, number, person, tense etc., and their agreement is needed. The particle is a class of words which include equivalents of English prepositions, conjunctions and adverbs etc. Particle has no inflection. (Adjective is formed by noun or verb.)

## 2 Noun unit

We classify noun roots into five types; common, proper, material, abstract, agentive. The common, material and abstract noun roots can be used as adjectives with no suffixes. Agentive comes from verb root.

We formalized attributes of noun units into following five tuples.
(gender, number, definiteness, case, conjunctive)

gender := masculine(M)/feminine(F)/common gen.(Cg)
number := singular(S)/plural(P)/common num.(Cn)
definite.:= definite(D)/indefinite(I)/undecided(U)
case := nominative(N)/accusative(A)/vocative(V) instrumentive(In)/ auxiliary(Au)/dative(Da)/ ablative(Ab)/ genitive(G)/locative(L)
conjun. := with suffix "Th"(Y)/without "Th"(N)

Example: Noun unit with the root GaS (tree).

| suf1 | suf2 | suf3 | noun unit | | attributes | | | | |
|---|---|---|---|---|---|---|---|---|---|
| - | - | - | GaS | (-s) | M | P | U | N/A/V | N |
| A | - | - | GaSa | (the -) | M | S | D | N/A/V | N |
| AK | - | - | GaSaK | (a -) | M | S | I | N/A | N |
| A | Ta | - | GaSaTa | (to the -) | M | S | D | Da | N |
| - | E* | - | GaSe* | (in -s ) | M | S | D | Au | N |
| - | WaLa | - | GaSWaLa | (on -s) | M | P | U | L | N |
| - | WaLaTa | | GaSWaLaTa | (to -s) | M | P | U | Da | N |
| - | EHi | - | GaSeHi | (in -s) | M | S | D | G | N |
| - | - | Th | GaSuTh | (-s and) | M | P | U | N/A | Y |
| AK | - | Th | GaSaKuTh | (a -, too) | M | S | U | N/A | Y |
| A | - | Th | GaSaTh | (- and) | M | S | D | N/A | Y |
| AK | - | Th | GaSaKaTh | (a - and) | M | S | I | A | Y |
| AK | IN | - | GaSaKiN | (from a -) | M | S | I | Au | N |
| - | WaLa | Th | GaSWaLaTh | (in -s too) | M | P | U | L | Y |
| - | E* | Th | GaSe*Th | (in -s too) | M | S | D | Au | Y |

We classify suffixes into three types; suf1, suf2, suf3.

**suf1** changes or adds gender, number and definiteness to the noun root. And it also indicates some cases (N, A, V).

*Gender:* masculine, feminine, common. Common gender root is such as KoTi (tiger), which becomes masculine with suffix A* (KoTiYa*) and feminine with suffix DeNa (KoTiDeNa). The root KoTi by itself represents tigers (plural) without regarding the gender.

*Number:* singular, plural, common.
Singular root can be changed into plural by some suffixes such as WaL, U.
RaTa + WaL = RaTaWaL (countries)
Plural root can be changed into singular by some suffixes such as A, A*.
GaS + A = GaSa (*tree*)
Common root can be changed into singular or plural by some suffixes such as A or O*.
KoTi + A = KoTiYa* (*tiger*)
KoTi + O* = KoTiYo* (*tigers*)
The root KoTi by itself also represents plural.

*Definiteness:* It is represented by such as EK, AK, A* .
MiNiS+ A* = MiNiSa* (*the man*)

MiNiS+ EK = MiNiSeK (*a man*)
GazNi+ AK = GazNiYaK (*a woman*)
GaS+ AK = GaSaK (*a tree*)

---

**Formalization of Sinhalese Morphology**

S. Herath‡, T.Ikeda, S.Ishizaki, Y.Anzai‡
‡Keio University, Yokohama 223
Electrotechnical Laboratory, Tsukuba-shi 305

*Case:* The suffixes such as A*, U, UN indicates cases; nominative, accusative and vocative(singular).

MiNiS + A* = MiNiSa*(the man) (N/A/V)
MiNiS + HU = MiNISSu(men) (N)
MiNiS + UN = MiNISSuN(men) (A)

**Suffix 2** represents the cases; dative, ablative, genitive, vocative(plural).

MiNISa* + Ta = MiNiSa*Ta(to the man) (Da)
MiNiSa* + GeN= MiNiSa*GeN(from the man) (Ab)
MiNiSa* + Ge* = MiNiSa*Ge*(man's) (G)
MiNiS + UNi = MiNiSuNi(men!) (V)

(note: A neutral word has no vocative form.)
(note: The remaining 3 cases, instrumental, auxiliary and locative are decided by particle units such as WiSiN for instrumental, NiSa* for auxiliary, KeReHi for locative.)

**suffix 3** is the conjunctive "Th", equivalent "too"and "and" in English.
(note: Other conjunctives are particle units such as SaHa (and), Ha* (and), Ho*(or).)

## 3 Verb unit

Verb unit consists of a root and a suffix. The suffix indicates various attributes of the verb unit. We have classified verb unit into 6 forms; participle, imperative, subjunctive, gerund, adjective and general form.

| | root | suffix | verb unit |
|---|---|---|---|
| Pre. participle | BaLa + | MiN | BaLaMiN (looking) |
| Past participle | BaLa + | La | BaLaLa(after looking) |
| Imperative | BaLa + | Nu | BaLaNu(look) |
| Subjunctive | BaLa + | WoTh | BaLuWoTh (if looks) |
| Gerund | BaLa + | I*Ma | BazLi*Ma (looking) |
| Adjective | BaLa + | Na | BaLaNa (looking) |

We formalised attributes of verb unit general form into six tuples.
(causative, voice, tense, person, gender, number).

causative := causative(C)/non-causative(Nc)
voice := active(A)/passive(P)
tense := present(Pr)/past(Pa)/future(Fu)
person := first(1)/second(2)/third(3)
gender := masculine(M)/feminine(F)
number := singular(Si)/plural(Pl)

Ka + Mu = KaMu(eat) [Nc, A, Pr, 1, (M/F), P]

At linking with suffixes, the verb root takes several different forms.

KaRa + INi = KeRiNi(happened)

We formalized these changes as 11 inflections including the real verb root.

KaRa := (KaRa KaRaz KeLe KeLe* KaRa KaRa KaRa KaRa KeRu KeRi KeRe)

In the suffix dictionary, we give informations on which inflection form to be linked with the suffix.

## 4 Linking:

We have to consider the linking factor when analysing noun units. There are 12 linking patterns with sub-linking patterns in general. We formalised them as linking rules by using the last and the first letter (of our romanized writing system) of the components linked.

Depending on the consonant(C) or the vowel(V) of the last letter in the component and the first letter of its subsequent, we have derived the following linking rules.

1 V + C => no change.
RaTa + WaL= RaTaWaL (countries)
2 C + V => V changes to lower case letter.
GaS + A => GaSa (a tree)
3 V + V => There are four situations. We give one here.
If the former V is other than a*/u*/e*/i/u then it is eliminated.
else .....
MoNaRa + A* => MoNaR + A*
=> MoNaRa*(the peacock).. by rule 2.
4.C + C => if the latter C is "H",
then eliminate it and duplicate the former C.
else no change.
MiNiS + Hu => (MiNiS + u =>) MiNiSSu (men)
GaS + WaLa=> GaSWaLa (in trees)

## 5 Discussion and further work

We have formalized Sinhalese morphology. Now we are implementing it on the computer as a part of Sinhalese syntax analyzer.
Sinhalese word order is more flexible than Japanese. We study its syntax formalization, parsing system and semantic analysis as the next step.

### Acknowledgment

We wish to acknowledge with thanks the support and encouragement from Natural Language Section, Electrotechnical Laboratory, and Anzai Laboratory, Keio University.

### References

[1] S.Herath, T. Ikeda, S. Yokoyama, H. Isahara and S. Ishizaki, " Sinhalese Morphological Analysis: A Step Towards Machine Processing of Sinhalese", Proceedings of Tools for Artificial Intelligence, TAI'89, October, 1989.

[2] Theodore G. Perera, " Sinhala Bhashawa", M.D. Gunasena (Sri Lanka), 1985.

[3] Kumaratunge Munidasa, "Viyakarana Vivaranaya", M.D. Gunasena (Sri Lanka), 1983.