Technical Reports in Language Technology
Report No 9, 2010
http://www.njas.helsinki.fi/salama

Language learning system using language analysis and disambiguation

Arvi Hurskainen
Department of World Cultures, Box 59
FIN-00014 University of Helsinki, Finland
arvi.hurskainen@helsinki.fi

Abstract

This report describes a language learning system, where a comprehensive language analysis system, including morphological analysis and disambiguation, is used as a basis for constructing various learning modules. One advantage of this approach is that it can be applied to learning systems with any type of morphological analyzer, including those that do not allow a space as part of input string.

Key Words: *language learning, morphological analysis, disambiguation*

1 Introduction

Earlier I reported on a language learning system that makes use of morphological analysis only (Hurskainen 2009). In that system, each word-form, when analyzed, produces all legal interpretations of the word-form. As a result, phrases with several words tend to produce a large number of readings. These readings are disambiguated using concordance rules and word order rules for controlling the correctness of the expression. Tests showed that it is possible to construct accurate learning modules even without a proper disambiguation component. The system was constructed using the Xerox fst package for morphological analysis (Karttunen and Beesley 2003). The characteristic feature of the Xerox system is that it allows the space as part of the input string. This facility can be used in constructing the morphological lexicon, so that legal sequences of words are allowed, and illegal ones blocked.

In this report I will describe an approach, where each word form is analysed separately regardless its context, and then the result is subjected to comprehensive disambiguation. This approach gives at least three advantages compared with the earlier approach. First, the morphological analysis system need not cross word boundaries and thus the danger of massive over-generation is avoided. Second, the already disambiguated strings are easy to handle in constructing learning modules. Third, the system can be implemented using any approach for constructing the morphological analyzer, because the control of word concatenation comes after the morphological analysis.

The basic linguistic analysis in this new system is performed using SWATWOL and SWACG, both of which are components in a language management system termed Swahili Language Manager (SALAMA[1]).

SWATWOL is a morphological analyzer of Swahili, and it makes use of finite-state methods including a two-level description (Koskenniemi 1983; Hurskainen 1992).

SWACG is the disambiguation module based on the theory of Constraint Grammar (Tapanainen 1996, 1999). The module performs morphological disambiguation and syntactic mapping. It also isolates multi-word expressions, a facility necessary in dictionary compilation and machine translation.

2 Inclusion of the learner's language

When learning a foreign language, it is important that the learner immediately knows what the words mean and how they are composed. The full disambiguated morphological analysis gives accurate information on the learning material, including glosses in the language of the learner.

Consider a simple sentence in (1), analysed with SWATWOL and disambiguated with SWACG.

(1)
```
"<<s>>"
      "<s>" { <s> }
"<mtoto>"
      "mtoto" N 1/2-SG { the } { child } CAP
"<anasoma>"
      "soma" V 1-SG3-SP VFIN { he } PR:na z [soma] { read } SVO
"<<\s>>"
      "<\s>" { <\s> }
```

It contains the surface text, the base form of each token, morphological information of each morpheme, and a gloss in English.

This can be formatted to better suit for learning purposes. In (2) the sentence is put on one line, and the tokens are collected in front of the analysis part.

(2)
```
mtoto anasoma     "mtoto"+N+1/2-SG{the}{child} "soma"+V+1/2-SG3-
SP+VFIN{he}+PR:na[soma]{read}
```

The example in (2) contains the lemma of each word. To make the result simpler, the lemma can be removed as in (3).

---

[1] SALAMA (Hurskainen 2004) is a package of programs for manipulating Swahili text. SWATWOL and SWACG are key components in this package. In addition to analysis and disambiguation, SALAMA includes modules for syntactic mapping, dictionary compilation and machine translation from Swahili to English.

(3)
```
 mtoto anasoma    +N+1/2-SG{the}{child} +V+1-SG3-
SP+VFIN{he}+PR:na[soma]{read}
```

In the next phase the rules check the word order and, if it is correct, the code is inserted in the end of the string (4). The code tells that the word order is noun plus verb and that it is correct.

(4)
```
 mtoto anasoma    +N+1/2-SG{the}{child} +V+1-SG3-
SP+VFIN{he}+PR:na[soma]{read} N+V_WO
```

If the word order is not correct, we get an output as in (5).

(5)
```
huyu wangu  +DEM+1-SG{this} +POSS+3-SG+SG1{my} DEM+POSS_!WO
```

In the next phase we check whether the concordance is correct. Here again, we use rules. If the concordance is correct, a code is inserted at the end of the string. The code CONC2 in (6) tells that the concordance is correct and that the construction has two words.

(6)
```
 mtoto anasoma    +N+1/2-SG{the}{child} +V+1-SG3-
SP+VFIN{he}+PR:na[soma]{read} N+V_WO CONC2
```

If the concordance is not correct, we get the output as in (7). The absence of the concordance mark is an indication of wrong concordance.

(7)
```
mtoto wanasoma    +N+1/2-SG{child} +V+2-PL3-
SP+VFIN{they}+PR:na[soma]{read} N+V_WO
```

3 Reporting

When each type of success or mistake is marked at the end of the output string, it is possible, using just these codes, to write rules for giving feedback to the learner. Examples of feedback follow.

(8)
```
 anasoma     +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} V
```
This is a correct Swahili word.
What do you like to learn?
If you want guided practising of various structures, type ***mtoto****.*
If you want to use your own vocabulary, please go ahead.

By typing any correct single Swahili word the learner gets the report as in (8) above. To see whether this is true with other words, we can make tests as in (9-11).

(9)
```
 tulipowatembelea +V+2-PL1-SP+VFIN{we}+PAST+16-SG-REL{when}+2-PL3-
OBJ+OBJ{them}[tembea]{walk} V
```
This is a correct Swahili word.
What do you like to learn?
*If you want guided practising of various structures, type **mtoto**.*
If you want to use your own vocabulary, please go ahead.

(10)
```
 utandawazi +N+11-SG{globalization} N
```
This is a correct Swahili word.
What do you like to learn?
*If you want guided practising of various structures, type **mtoto**.*
If you want to use your own vocabulary, please go ahead.

(11)
```
 kitu +N+7/8-SG{thing} N
```
This is a correct Swahili word.
What do you like to learn?
*If you want guided practising of various structures, type **mtoto**.*
If you want to use your own vocabulary, please go ahead.

3.1 Reporting on word order

Next we will study the feedback in word order problems. Consider the examples in (12-15).

(12)
```
 wangu mtoto      +POSS+3-SG+SG1{my} +N+1/2-SG{child}
POSS+N_!WOINIT
```
Possessive pronoun cannot initiate a phrase!

(13)
```
 mtoto huyu wangu +N+1/2-SG{child} +DEM+1-SG{this} +POSS+3-
SG+SG1{my} DEM+POSS_!WO
```
Demonstrative pronoun cannot be before a possessive pronoun!

(14)
```
 mtoto mmoja huyu +N+1/2-SG{child} +NUM+1-SG+NUM-INFL{one} +DEM+1-
SG{this} NUM+DEM_!WO
```
Numeral cannot be before a demonstrative pronoun!

(15)
```
 mtoto mmoja wangu      +N+1/2-SG{child} +NUM+1-SG+NUM-INFL{one}
+POSS+3-SG+SG1{my} NUM+POSS_!WO
```
Numeral cannot be before a possessive pronoun!

When there are more than one word order mistake, they are treated one by one (16-17).

(16)
```
 mtoto mmoja mzuri wangu      +N+1/2-SG{child} +NUM+1-SG+NUM-
INFL{one} +A-INFL+1-SG{good} +POSS+3-SG+SG1{my} A+POSS_!WO
```
Adjective cannot be before a possessive pronoun!

Now we correct that string and get as in (17).

(17)
```
 mtoto mmoja wangu mzuri      +N+1/2-SG{child} +NUM+1-SG+NUM-
INFL{one} +POSS+3-SG+SG1{my} +A-INFL+1-SG{good} NUM+POSS_!WO
```
Numeral cannot be before a possessive pronoun!

Then we have to correct the input string as advised (18).

(18)
```
 mtoto wangu mmoja mzuri      +N+1/2-SG{child} +POSS+1-SG+SG1{my}
+NUM+1-SG+NUM-INFL{one} +A-INFL+1-SG{good} NUM+A_!WO
```
Numeral cannot be before an adjective!

When we make still one more adjustment, we get the right word order (19)

(19)
```
 mtoto wangu mzuri mmoja      +N+1/2-SG{child} +POSS+1-SG+SG1{my}
+A-INFL+1-SG{good} +NUM+1-SG+NUM-INFL{one} N+POSS+A+NUM_WO CONC4
```
Word order and concordance are correct!

3.2 Reporting on concordance

Swahili has a total of 15 noun classes, and constituents depending on the noun have to agree with it. In the current application, each noun class is marked using a number from one to eighteen (classes 12, 13 and 14 are not in use). Writing rules for normal cases is fairly easy, because one has just to make sure that each constituent contains the same class number. In (20) is a bit longer example.

(20)
```
 mtoto wangu mzuri huyu mmoja anasoma     +N+1/2-SG{child}
+POSS+1-SG+SG1{my} +A-INFL+1-SG{good} +DEM+1-SG{this} +NUM+1-
SG+NUM-INFL{one} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read}
N+POSS+A+DEM+NUM+V_WO CONC6
```
Word order and concordance are correct!

If there is a mistake in concordance, we get an output as in (21).

(21)
mtoto zangu anasoma +N+1/2-SG{child} +POSS+10-PL+SG1{my} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} **N+POSS+V_WO**
Word order is correct but concordance is not!

4 Control of word order precedes the control of concordance

It can also happen that the word order and concordance are wrong. Intuitively one would start with correcting the word order. In the current implementation, the learning system instructs the learner first to correct the word order and then correct the concordance. The examples in (22-24) illuminate this.

(22)
zangu mtoto anasoma +POSS+10-PL+SG1{my} +N+1/2-SG{child} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} **POSS+N_!WOINIT**
Possessive pronoun cannot initiate a phrase!

When the word order is corrected, the system asks to correct the concordance (23).

(23)
mtoto zangu anasoma +N+1/2-SG{child} +POSS+10-PL+SG1{my} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} **N+POSS+V_WO**
Word order is correct but concordance is not!

When the concordance is corrected, the string is accepted (24).

(24)
mtoto wangu anasoma +N+1/2-SG{child} +POSS+1-SG+SG1{my} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} **N+POSS+V_WO CONC3**
Word order and concordance are correct!

5 Alternative word order

The standard word order in Swahili is: Noun+Poss+Adj+Dem+Num+Verb. Above we have seen warnings about the wrong word order. However, the demonstrative pronoun may also appear in front of the noun (25-26).

(25)
huyu mtoto anasoma +DEM+1-SG{this} +N+1/2-SG{child} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} **DEM+N+V_WO CONC3**
Word order and concordance are correct!

(26)
huyu mtoto wangu mzuri mmoja anasoma +DEM+1-SG{this} +N+1/2-SG{child} +POSS+1-SG+SG1{my} +A-INFL+1-SG{good} +NUM+1-SG+NUM-INFL{one} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read}
DEM+N+POSS+A+NUM+V_WO CONC6
Word order and concordance are correct!

6 Irregular concordance

Although the concordance sounds simple, its full mastery requires a lot of training. A number of factors contribute to this. First, the surface morpheme sets for various constituents are different. For example, in (26) above we see that although each constituent inflects according to class 1, the class markers in words have four different types of markers, i.e. *-yu*, *m-*, *w-*, and *a-*. When each of the 15 noun classes has its own surface forms for each word category, it is not easy for the learner to master all these morpheme sets. Second, animate nouns that do not belong to the class 1/2 require various deviant inflection patterns. Such nouns are found in classes 7/8, 9/6 and 9/10. When the formal class of the noun, defined by its surface form, clashes with its semantic meaning, the resulting concordance pattern is a mixture of classes. Ambiguity of the nouns adds to the complexity even more. Examples in (27-28) illuminate this.

(27)
kifaru wetu mkubwa huyu anakimbia +N+7/8-SG{rhinoceros} +POSS+1-SG+PL1{our} +A-INFL+1-SG{big} +DEM+1-SG{this} +V+1-SG3-SP+VFIN{he}+PR:na[kimbia]{run} **N+POSS+A+DEM+V_WO CONC5**
Word order and concordance are correct!

(28)
kifaru chetu kikubwa hiki kinakimbia +N+7/8-SG{military_tank} +POSS+7-SG+PL1{our} +A-INFL+7-SG{big} +DEM+7-SG{this} +V+7-SG-SP+VFIN{it}+PR:na[kimbia]{run} **N+POSS+A+DEM+V_WO CONC5**
Word order and concordance are correct!

In (27) the noun *kifaru* (rhinoceros) has an irregular concordance, while *kifaru* (military tank) in (28) has a regular concordance.

There is also a small group of nouns meaning disabled persons that have a deviant concordance pattern (29-30).

(29)
kiwete wetu mdhaifu yule anatembea +N+7/8-SG{cripple} +POSS+1-SG+PL1{our} +A-INFL+1-SG{weak} +DEM+1-SG{that} +V+1-SG3-SP+VFIN{he}+PR:na[tembea]{walk} **N+POSS+A+DEM+V_WO CONC5**
Word order and concordance are correct!

(30)
viwete wetu wadhaifu wale wanatembea +N+7/8-PL{cripple} +POSS+2-PL+PL1{our} +A-INFL+2-PL{weak} +DEM+2-PL{those} +V+2-PL3-SP+VFIN{they}+PR:na[tembea]{walk} **N+POSS+A+DEM+V_WO CONC5**
Word order and concordance are correct!

There are a number of nouns meaning animates, animals as well as humans, that belong to classes 9/10 and 9/6. Sometimes the same noun is classified to both groups, and the concordance is irregular.

(31)
bibi yangu yule anasoma +N+9/6-SG{grandmother} +POSS+9-SG+SG1{my} +DEM+1-SG{that} +V+1-SG3-SP+VFIN{she}+PR:na[soma]{read} **N+POSS+DEM+V_WO CONC4**
Word order and concordance are correct!

The plural of *bibi* and *bwana* requires concordance from three different classes (32).

(32)
mabibi zetu hawa wanasoma +N+9/6-PL{grandmother} +POSS+10-PL+PL1{our} +DEM+2-PL{these} +V+2-PL3-SP+VFIN{they}+PR:na[soma]{read} **N+POSS+DEM+V_WO CONC4**
Word order and concordance are correct!

(33)
mabwana zetu hawa wanasoma +N+9/6-PL{lord}+MALE+HUM +POSS+10-PL+PL1{our} +DEM+2-PL{these} +V+2-PL3-SP+VFIN{they}+PR:na[soma]{read} **N+POSS+DEM+V_WO CONC4**
Word order and concordance are correct!

However, the singular of *bwana* has a strange concordance. If we try to use the same concordance pattern as with *bibi*, we get an error message (34).

(34)
bwana yangu huyu anasoma +N+9/6-SG{lord}+MALE+HUM +POSS+9-SG+SG1{my} +DEM+1-SG{this} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} **N+POSS+DEM+V_WO**
Word order is correct but concordance is not!

The correct concordance of *bwana* is in (35-36).

(35)
bwana wangu huyu anasoma +N+9/6-SG{lord}+MALE+HUM +POSS+1-SG+SG1{my} +DEM+1-SG{this} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read} **N+POSS+DEM+V_WO CONC4**
Word order and concordance are correct!

(36)
```
 huyu bwana wangu anasoma     +DEM+1-SG{this} +N+9/6-
SG{lord}+MALE+HUM +POSS+1-SG+SG1{my} +V+1-SG3-
SP+VFIN{he}+PR:na[soma]{read} DEM+N+POSS+V_WO CONC4
```
Word order and concordance are correct!

A few more nouns denoting relatives are in (37-39).

(37)
```
 kaka yangu mzuri anashughulika     +N+9/10-
SG{:elder_brother}+PERS +POSS+9-SG+SG1{my} +A-INFL+1-SG{good}
+V+1-SG3-SP+VFIN{he}+PR:na[shughulika]{be_busy} N+POSS+A+V_WO
CONC4
```
Word order and concordance are correct!

(38)
```
 kaka zangu wazuri wanashughulika   +N+9/10-
PL{:elder_brother}+PERS +POSS+10-PL+SG1{my} +A-INFL+2-PL{good}
+V+2-PL3-SP+VFIN{they}+PR:na[shughulika]{be_busy} N+POSS+A+V_WO
CONC4
```
Word order and concordance are correct!

(39)
```
 binti yangu mzuri anashughulika    +N+9/10-SG{daughter} +POSS+9-
SG+SG1{my} +A-INFL+1-SG{good} +V+1-SG3-
SP+VFIN{he}+PR:na[shughulika]{be_busy} N+POSS+A+V_WO CONC4
```
Word order and concordance are correct!

7 Inflecting and non-inflecting adjectives and numbers

Swahili adjectives and numbers fall into two categories according to their behaviour in context. Adjectives and numbers of Bantu origin do inflect, but those of Arabic origin do not. This traditional division is sometimes blurred, especially in case of new derived words. The problem occurs particularly in such neologisms that formally resemble inflecting adjectives but in fact have been derived from Arabic.

Inflecting adjectives and numbers behave as in (40-41).

(40)
```
 watoto wazuri watatu wanasoma      +N+1/2-PL{child} +A-INFL+2-
PL{good} +NUM+2-PL+NUM-INFL{three} +V+2-PL3-
SP+VFIN{they}+PR:na[soma]{read} N+A_WO CONC4
```
Word order and concordance are correct!

(41)
```
 kisu kizuri kimoja kimekata  +N+7/8-SG{knife} +A-INFL+7-SG{good}
+NUM+7-SG+NUM-INFL{one} +V+7-SG-SP+VFIN{it}+PERF:me[kata]{cut}
N+A_WO CONC4
```

Word order and concordance are correct!

Mistake in inflecting a non-inflecting adjective or number causes an error message (42-43).

(42)
```
 kisu kibutu kimoja kimekata  +N+7/8-SG{knife} Heur +NUM+7-SG+NUM-
INFL{one} +V+7-SG-SP+VFIN{it}+PERF:me[kata]{cut}
```
Please check spelling!

(43)
```
 visu vizuri visita vimekata  +N+7/8-PL{knife} +A-INFL+8-PL{good}
Heur +V+8-PL-SP+VFIN{they}+PERF:me[kata]{cut} N+A_WO
```
Please check spelling!

Note that the code Heur indicates the word that is misspelled. After correcting the spelling we get correct strings (44-45).

(44)
```
 kisu butu kimoja kimekata    +N+7/8-SG{knife} +A-UNINFL{blunt}
+NUM+7-SG+NUM-INFL{one} +V+7-SG-SP+VFIN{it}+PERF:me[kata]{cut}
N+A_WO CONC4
```
Word order and concordance are correct!

(45)
```
 visu vizuri sita vimekata    +N+7/8-PL{knife} +A-INFL+8-PL{good}
+NUM+NUM-UNINFL{six} +V+8-PL-SP+VFIN{they}+PERF:me[kata]{cut}
N+A_WO CONC4
```
Word order and concordance are correct!

8 Possibilities for various learning approaches

Above I have described the basic methods of how to make use of the comprehensive analysis system of the language in constructing language learning systems. Characteristic to this approach is that instead of operating with surface word-forms, the system makes use of the grammar of the language. Therefore, it is possible to construct rule systems on the basis of the grammar instead of surface phenomena. This has two major consequences. First, the scope of the learning material covers the whole lexicon, that is, any word of the language can be used in exercises. Second, the number of rules needed for controlling the learning process is manageable. One rule may cover hundreds, even thousands, of surface cases.

8.1 Learning from scratch

Learning a language through grammar is the safest way of learning it correctly. After mastering the grammar, the learner has a solid base for improving language skills by increasing the mastery of vocabulary. Reading word lists is dull and it is not a very intuitive learning method, if the learned word cannot be practised in context. The use in context is difficult, if there is no control on the correctness of the construction. The system described here gives the learner an interactive learning environment, where feedback and advice are given on each type of error - and correctness is approved.

What about if the learner is a full beginner? Can the learning system be constructed so that the learner gets off without external help? I think that access to a good grammar description is necessary for a learner in any case. The grammar can be a digital package that is part of the learning system. The learner would read it first to get a picture of the structure of the language. But it is also possible to construct learning modules that advice the learner by the try-and-error method. Let us demonstrate this approach using one guided tour to Swahili.

Let us assume that the learner knows at least one Swahili word, in any form. Consider responses in (46-48).

(46)
```
 jambo      +N+5/6-SG{matter} N
```
This is a correct Swahili word.
What do you like to learn?
*If you want guided practising of various structures, type **mtoto**.*
If you want to use your own vocabulary, please go ahead.

(47)
```
 waaminifu  +A-INFL+2-PL{faithful} A
```
This is a correct Swahili word.
What do you like to learn?
*If you want guided practising of various structures, type **mtoto**.*
If you want to use your own vocabulary, please go ahead.

(48)
```
 kilichonipendeza +V+7-SG-SP+VFIN{it}+PAST+7-SG-REL{which}+1-SG1-
OBJ+OBJ{me}[pendeza]{please} V
```
This is a correct Swahili word.
What do you like to learn?
*If you want guided practising of various structures, type **mtoto**.*
If you want to use your own vocabulary, please go ahead.

The response asks the learner to type the word *mtoto*. The learner does not know its meaning.

(49)
```
 mtoto      +N+1/2-SG{child} N-GUIDED
```
OK. Combine this noun with adjective 'zuri'!

To the word mtoto should be added the adjective *zuri*. We do not know whether it should be before or after *mtoto*, or what its form should be. We try (50).

(50)
```
 zuri mtoto +A-INFL+5-SG{good} +N+1/2-SG{child} A+N_!WOINIT
```
Adjective cannot initiate a phrase!

The word order was wrong. We change the order (51).

(51)
```
 mtoto zuri +N+1/2-SG{child} +A-INFL+5-SG{good} N+A_WO CONC2
```
Word order is correct but concordance is not!

Now the word order is correct, but the concordance is wrong. Another try, following the example of the noun (52).

(52)
```
 mtoto mzuri      +N+1/2-SG{child} +A-INFL+1-SG{good} N+A_WO CONC2
```
*OK. Add to this string a possessive pronoun **angu**!*

Now finally it is correct. Then we are asked to add the possessive pronoun *angu*. We try to put it to the end, because it seemed to work with the adjective. Let us also put *m-* in front of the word, because that was required for the adjective (53).

(53)
```
 mtoto mzuri mangu      +N+1/2-SG{child} +A-INFL+1-SG{good} Heur
```
Please check spelling!

Spelling did not go right for *angu*. Possessive pronouns have a different set for class markers. A look at concordance table helps to find the correct one (54).

(54)
```
 mtoto mzuri wangu      +N+1/2-SG{child} +A-INFL+1-SG{good}
+POSS+3-SG+SG1{my} A+POSS_!WO
```
Adjective cannot be before the possessive pronoun! It comes immediately after the noun.

Now after changing the word order we get as in (55).

(55)
```
 mtoto wangu mzuri      +N+1/2-SG{child} +POSS+1-SG+SG1{my} +A-
INFL+1-SG{good} N+POSS+A_WO CONC3
```
OK. Add to this string a numeral!

A numeral is added to the string. It is found in the vocabulary list. We do here as before (56).

(56)
```
 mtoto wangu mzuri moja +N+1/2-SG{child} +POSS+1-SG+SG1{my} +A-
INFL+1-SG{good} +NUM+5-SG+NUM-INFL{one} N+POSS+A+NUM_WO
```
Word order is correct but concordance is not!

Now we succeeded with the word order, but concordance has a problem. A look at the concordance table helps us (57)

(57)
```
 mtoto wangu mzuri mmoja      +N+1/2-SG{child} +POSS+1-SG+SG1{my}
+A-INFL+1-SG{good} +NUM+1-SG+NUM-INFL{one} N+POSS+A+NUM_WO CONC4
```
OK. Add to this string a demonstrative pronoun!

We look at the concordance table and select one of the appropriate pronouns. We type it after the previous string (58)

(58)
```
 mtoto wangu mzuri mmoja huyu +N+1/2-SG{child} +POSS+1-SG+SG1{my}
+A-INFL+1-SG{good} +NUM+1-SG+NUM-INFL{one} +DEM+1-SG{this}
NUM+DEM_!WO
```
Numeral cannot be before a demonstrative pronoun!

The pronoun was not in the correct place. We change the order as instructed (59).

(59)
```
 mtoto wangu mzuri huyu mmoja +N+1/2-SG{child} +POSS+1-SG+SG1{my}
+A-INFL+1-SG{good} +DEM+1-SG{this} +NUM+1-SG+NUM-INFL{one}
N+POSS+A+DEM+NUM_WO CONC5
```
*OK. Add to this string the verb **soma**!*

Now the pronoun is in the correct place and the verb *soma* should be added. It is obvious that it should be in a finite form. Let us assume that the learner, knowing the basics of verb structure, has figured out a construction that has the subject marker and present tense marker (60).

(60)
```
 mtoto wangu mzuri huyu mmoja mnasoma     +N+1/2-SG{child}
+POSS+1-SG+SG1{my} +A-INFL+1-SG{good} +DEM+1-SG{this} +NUM+1-
SG+NUM-INFL{one} +V+2-PL2-SP+VFIN{you}+PR:na[soma]{read}
N+POSS+A+DEM+NUM+V_WO
```
Word order is correct but concordance is not!

The prefix *m-* as the subject marker was not correct. Again, a look at the concordance table helps (61).

(61)
```
 mtoto wangu mzuri huyu mmoja anasoma     +N+1/2-SG{child}
+POSS+1-SG+SG1{my} +A-INFL+1-SG{good} +DEM+1-SG{this} +NUM+1-
SG+NUM-INFL{one} +V+1-SG3-SP+VFIN{he}+PR:na[soma]{read}
N+POSS+A+DEM+NUM+V_WO CONC6
```
OK. This is a full sentence with a long noun phrase. Now do the same exercise using plural forms. Continue by typing ***watoto****!*

The beginner has managed to formulate a fairly complex sentence with the additional help of a concordance table and vocabulary list only. The learning system reports on mistakes in this order: spelling > word order > concordance, and gives advise when needed. The formulation of plural forms must be much easier, because the word order is already known.

The tour could continue through all noun classes as one lesson, although this is hardly advisable. Therefore, it is better to cut the tour into separate lessons. How to inform the learner what the lessons are and how to start a particular lesson? One method is to inform about these lessons in the opening feedback. That is, when the learner types a Swahili word, the output looks like in (62).

(62)
```
 wangu       +POSS+3-SG+SG1{my} POSS
```
This is a correct Swahili word.
What do you like to learn?
If you want guided practising of various structures
type one of the following words:
mtoto, mti, funzo, kiti, saa, ufagio, bibi,
ndugu, kifaru, kiziwi, mahali, kusoma*.*
If you want to use your own vocabulary, please go ahead.

The learner gets a list of nouns, each of which initiates a lesson of that particular noun type. For the introductory tour, these lessons can be tied together, so that, by following the instructions, the learner goes through all regular and irregular concordance types. For example, if the learner types the word *kiti*, the lesson starts from that word and continues through *saa, ufagio, bibi, ndugu, kifaru, kiziwi, mahali*, and *kusoma*, to the end. The learner can 'jump in' on any word and continue from that point.

In order to properly train constructions of each noun type, specialised training sessions can be constructed. Such guided training may include several nouns of the same type, as well as several words from other word categories, i.e. possessive pronouns, adjectives, demonstrative pronouns, numerals, and verbs. There are plenty of possibilities for constructing many types of interactive lessons with detailed guidance.

8.2 Comprehensive learning system

Guided learning tours are useful for beginners, but the real power of the system lies in its comprehensiveness. Tens of thousands of words and millions of word-forms are covered.

While language learning materials with no access to language analysis are restricted to a series of tutorials and exercises, the system described above allows for several types of self-learning possibilities. For example, vocabulary learning from a vocabulary list is dull and counter-intuitive. In this system it can be combined with exercises in context, and the system reports on mistakes and guides the learner. Thus each POS category can be learned in real context without being restricted to a narrow set of vocabulary.

Because the system is comprehensive and accurate in terms of vocabulary and grammar, it serves also as a grammar checker. Even experienced writers would benefit from the system.

9 Considerations on user interfaces

Language learning environments should be made accessible to the target groups through various media. In terms of terminals, that is, the devices that the learners would use in learning, the development is rapid. The distinction between computers and hand phones is narrowing down, although not totally disappearing. Both serve as terminals for increasing numbers of applications. It is self-evident that a computer with a proper keyboard would be suitable as a learning device. But what is the case with hand phones? They are portable and their distribution is much larger than that of computers. Because of their portability they would be ideal for language learning. The bottleneck is not the memory needed by the system, but rather the insufficient convenience in typing text. Also in this respect the situation is changing rapidly. Hand phones with a proper keyboard have come to the market. And because the need of typing in language learning is quite limited, even small keyboards would suit to the purpose.

The access to the learning system can be arranged either by downloading the system to the terminal or by using the system over the Internet. Which one is better in each case depends on factors such as the capacity of the terminal, reliability of the Web, and the pricing policy. In any case, bright future can be expected for intelligent language learning systems.

10 Conclusions

This report has demonstrated that an accurate and comprehensive language learning system can be constructed on the basis of a morphological analyzer and disambiguator. Although similar modules can be constructed on the basis of the morphological analyzer alone, without accompanying disambiguator, the use of a proper disambiguator makes the development of the learning system easier. However, the use of a disambiguator is not without dangers, because if the disambiguator makes a mistake, it is not possible to trace the correct choice afterwards. Yet this is rather an implementational than a technical problem.

References

Beesley K. and Karttunen L. 2003. *Finite State Morphology*. Series: Studies in Computational Linguistics, 3. Stanford: CSLI Publications.

Hurskainen A. 1992. "A Two-Level Computer Formalism for the Analysis of Bantu Morphology. An Application to Swahili". *Nordic Journal of African Studies* 1(1): 87-122. http://www.njas.helsinki.fi

Hurskainen, A. 2004. "Swahili Language Manager: A Storehouse for Developing Multiple Computational Applications". *Nordic Journal of African Studies* 13(3): 363–397. http://www.njas.helsinki.fi

Hurskainen, A. 2009. *Intelligent Computer-Assisted Language Learning: Implementation to Swahili*. Technical Reports in Language Technology. Report No 3, 2009. www.njas.helsinki.fi/salama/technical-reports.html

Koskenniemi, K. 1983. *Two-level morphology: a general computational model for word-form recognition and production*, Publications No. 11. University of Helsinki: Department of General Linguistics.

Tapanainen, P. 1996. *The Constraint Grammar Parser CG-2.* Publications No. 27. Department of General Linguistics, University of Helsinki.

Tapanainen, P. 1999. *Parsing in two frameworks: finite-state and functional dependency grammar*. Ph.D. thesis, Department of General Linguistics, University of Helsinki.