Technical Reports in Language Technology
Report No 5, 2009
http://www.njas.helsinki.fi/salama

Machine Translation of the Bible

Arvi Hurskainen
Institute for Asian and African Studies, Box 59
FIN-00014 University of Helsinki, Finland
arvi.hurskainen@helsinki.fi

Abstract

The main trend in machine translation (MT) continues to make use of parallel corpora and various machine learning techniques, while the rule-based approaches have remained as a small niche in the field. Although the value of rule-based MT has not been denied, it is considered too costly and labour-intensive to become a real alternative in the field. However, there are application areas, where statistical machine translation is not possible, because the basic resources for doing it are missing. One such area is translation into a language that does not have texts, and has hardly a writing system. The paper describes a rule-based system for accomplishing this, applying the approach to Swahili to Luganda machine translation of the Bible.

Key Words: machine translation, rule-based machine translation, symbolic machine translation, Bible translation

Abbreviations

CG	Constraint Grammar
MT	machine translation
MWE	multiword expression
SALAMA	Swahili Language Manager
SL	source language
TL	target language

1 Introduction

Hundreds of man years are used every year in translating Bible texts to a large number of languages. A number of computational tools have been developed for assisting in this tedious work, and with the help of these tools it has been possible to speed up some phases in translation and correction. The translation of the whole Bible into a new language often takes more than ten years to complete. Therefore, it is justified to ask whether it is possible to speed up the work by developing such tools that automate at least part of the job.

It is not realistic to think that a computer can make faultless translation of all parts of the Bible text. But although fully acceptable translation cannot be achieved, there are many areas in translation that can quite safely be made automatically.

In this paper, main phases and problems in Bible translation are discussed, and a method for constructing a MT system from one language to another is described. The method was implemented for translation from Swahili to Luganda, both of which are Bantu languages.

2 Motivation of the attempt

Because there already was a MT system from Swahili to English, developed as part of the SALAMA tool package (Hurskainen 2004b), some of my colleagues were wondering whether it could be applied also to Bible translation. Although I first was somewhat hesitant about its feasibility for this kind of task, I promised to investigate the matter. I decided to apply the SALAMA system for translating from Swahili to Luganda, because I had access to necessary resources in that language[1]. The input would be the Bible in Swahili, and the output would be its translation into Luganda.

I was also interested to see, how a MT system developed to one target language could be adjusted to another target language, and how much work it would require. Bible text, although it is domain-specific, contains many kinds of styles. Nevertheless, almost all of it is standard prose text and therefore suitable for MT.

Failures in MT are often due to mistakes and inconsistencies in the source language. The question has bee raised whether source text writers should pay more attention to the translatability of the text (Bernth and Gdanie 2002) to improve translation quality. In the case of the Bible, the text is exceptionally clean and grammatical. Therefore, the problems haunting in news texts should not be prominent in translating the Bible.

3 Rule-based method – the only alternative

The large majority of MT systems developed currently do not make extensive use of linguistic knowledge. These approaches are based on machine learning, parallel corpora, probabilities, translation examples, and occasionally, in trying to improve performance, they try to include also linguistic knowledge into the system (Groves and Way 2006).

As a result we currently have ‘machine translation’ systems for translating Web pages between various languages. It is no wonder that most Web users laugh at the translation results, and avoid everything that hints at MT. The current trend in the field has not changed much, and disillusionment continues and perhaps even deepens. The recent publications in the field, for example in the leading periodical Machine Translation, do not promise major advances in the field. However, there are some exceptions, such as the inclusion of ‘grammaticality’ into statistical MT (Riezler and Maxwell 2006). Ten Hacken (2001) saw even a ‘revolution’ to have taken place in the field.

---

[1] The necessary resources include the transfer lexicon from English ton Luganda and the grammar of Luganda. While working in 2008 for four months as a Visiting Professor at Makerere University, I conducted an intensive course on language technology. One of the languages, for which we constructed morphological analyzers, was Luganda. This exercise helped us, in fact forced us, to work out a detailed description of the language.

Yet there is no doubt that MT is an extremely important field, and there is no reason to doubt that one day the languages of the world can be translated to other languages in the way that is acceptable.

When thinking of MT to minority languages, such as the languages of Africa, the main approach in MT is not applicable at all, because all necessary resources are missing. If some national languages of the European Union are regarded as low resource languages (Carl et al. 2008), how much more this is true of African languages. This is perhaps not only a regrettable fact, because it excludes the use of the main stream methods, and forces to use the other alternative, that is, the rule-based approach.

I have no reason to argue which method is superb, because the other alternative is excluded. Nevertheless, I fail to understand why we do not make use of all available linguistic information in MT, because linguistic analysis makes explicit such features that can be used in formulating rules for guiding the processes.

A MT system that is composed of several sequentally ordered modules is also easy to maintain and correct (Nyberg et al. 1994). And the most important factor in choosing a rule-based system is that it is the only approach suitable for the purposes discussed here. One could think that parallel corpora could be used in devising a MT system between two minority languages, for example by using Bible translations of the two languages as a parallel corpus. Well, but if there is no text at all in the target language, perhaps not even a writing system, how can we use statistical methods?

The persistent problems in statistical MT include finding translations to low-frequency words (Pekar et al. 2006), because they do not appear in the corpus, or they appear there too rarely for training.

4 Assumptions

The ideal in translation, including Bible translation, is that it should be performed from the original source language to the target language. In the case of the Bible it would mean Hebrew and Greek. However, only occasionally the translation is carried out from the original source text, because of a number of reasons. In many translation teams the knowledge of original languages is poor or missing, and the translators have to rely on translations in languages that they know.

There might also be some advantage of translating from a language, whose language structure is similar to the structure of the target language. The speakers of these two languages also share many cultural features, and it could be expected that the culture-specific translation solutions made in translating to one language could be made also in translating into another related language. This is analogous to the saying, 'Why to invent the wheel twice?' In fact, there are hundreds of translation teams pondering on the same questions, and coming more or less to the same conclusions. Why to do the same over and over again? A MT system can give also culture-specific translations, not only general standard translations.

It is also assumed that translation between related languages is less difficult than translation between unrelated and structurally very different languages. This means that

translating from Swahili to Luganda would be easier than from Swahili to English. This will be demonstrated below, especially when dealing with the problem of re-ordering constituents[2].

The vocabulary problems in translation vary tremendously. The Bible contains names of animals and plants as well as environment-specific concepts that hardly have a ready translation in the target language. In these cases the MT cannot help, but it can register precisely each of such problematic translations.

On the other hand, the Bible contains about 2750 proper names, which could be safely translated in all places where they occur. Also the time-consuming translation of long chronologies can be safely translated with MT.

5 Approach and phases in translation

Here is a brief description of the approach used in translating the Bible from Swahili into Luganda. Although the approach is described using two concrete Bantu languages as examples, the solutions made for solving problems are not language-specific. After the general description in this chapter, a more detailed description is given in Chapter 6, with snapshots of various intermediate phases of the translation process.

5.1 Decomposing the text

The rule-based MT makes use of two main components of language description, the lexicon and the grammar. The lexicon contains, in addition to the stem lexemes, also all other morphemes, such as prefixes and suffixes. These affixes may have inflectional or derivational roles. The grammar defines how the affixes can be attached to the stem. The grammar also contains rules for morpho-phonological variation, as well as for variations required by concordance rules.

In other words, the surface language is represented also on another level, that is, on the level where linguistic features are classified. This representation makes it possible to treat linguistic features as sets rather than as individual strings. This gives a tremendous advantage compared with statistical translation that operates with surface strings. The advantage is even more prominent in languages with rich morphology, such as Bantu languages, where a single verb may have millions of formally correct forms.

Processing the languages into the fully analyzed format requires (a) tokenization, (b) morphological analysis, (c) morphological disambiguation, (d) isolation of multiword-expressions, (e) syntactic mapping, and (f) semantic disambiguation. This is the format of the source language, from which the translation into the target language begins.

[2] By constituents I mean individual words and also such morphemes that constitute part of a word. In MT, the order of these constituents needs to be controlled to meet the requirements of the target language.

5.2 Lexical transfer to target language

The composition of the TL starts from transferring constituents of the SL into the format required by the TL. The first step is the lexical transfer, which includes the transfer of stems and possibly also the transfer of other lexical constituents, such as prefixes and suffixes. It depends on the structure of the TL which constituents are relevant for transfer. For example, in the case of Swahili-to-English translation, only stems can be transferred, because English hardly has any morphology. In the case of Swahili-to-Luganda translation, a large number of morphological elements other than stems can be transferred. Examples below will illustrate this point.

However, also on the level of stem transfer there are problems in finding the most appropriate equivalent in the TL. Mel'čuk and Wanner (2001) have elaborated on this issue in detail.

The lexical transfer to the TL can take place through an intermediate language, inter lingua, or directly from the SL to the TL. In the case discussed in this paper, the direct transfer would be most appropriate for a number of reasons. Swahili and Luganda are related languages, and in many cases the mapping from the SL to the TL could be done directly.

Now, because no transfer lexicon exists between these languages, the transfer must be done via English. It is a matter of taste whether English in this case should be termed as an inter lingua or not. But the fact is that English in between unnecessarily complicates the translation process.

5.3 Composing the surface form of target language

As was said above, the transfer from the SL to the TL (via English in this case) takes place only on the lexical level. In addition to transferring the lexical stem, also information on the noun group affiliation of nouns is marked, because the noun group may be different in the SL and TL. The grammatical information is stored in various grammatical tags inherited from the SL.

The surface form of the TL is constructed using the transferred lexical stems and the grammatical information stored in tags. Because both the SL and the TL are Bantu languages, often the conversion is unproblematic. However, there are a number of differences on the surface of these two languages. Such differences include different noun group affiliation, differences in singular/plural expressions, and differences in constructing locative and relative expressions. Also several verb-forms are constructed differently, which requires individual treatment of each type of verb-forms. Particularly problematic is the production of the forms of the subject prefix and object prefix of the verb, the forms of which depend on the noun class of the subject and object respectively.

The solution for handling the divergent realizations of structures in the TL is to introduce tags also from the target language in cases where the tags of the SL and TL are in conflict. When such a conflict occurs, the surface form is constructed using the tag(s) of the TL. Otherwise the tag(s) of the SL are used.

6 Detailed description of phases in translation

Below is a description of the translation procedure step by step from the source text in Swahili into the target text in Luganda. The example text is from Matthew 28:18-20. For the sake of space restrictions, in some examples only part of text is used.

6.1 Formatting the Bible text in source language

In order to handle safely the source language in translation, the source text is pre-formatted. Each verse is given a unique code, from which it can be identified in all phases of translation. The verse, regardless its length, is treated as a single unit (1).

(1)
[MAT28_18/] Yesu akaja kwao, akasema nao, akawaambia, Nimepewa mamlaka yote mbinguni na duniani.

[MAT28_19/] Basi, enendeni, mkawafanye mataifa yote kuwa wanafunzi, mkiwabatiza kwa jina la Baba, na la Mwana, na la Roho Mtakatifu;

[MAT28_20/] na kuwafundisha kuyashika yote niliyowaamuru ninyi; na tazama, mimi nipo pamoja nanyi siku zote, hata ukamilifu wa dahari.

6.2 Tokenization

In tokenization, the source text is formatted to meet the requirements of the analysis process. This includes the transformation of capital letters temporarily to lower case, detaching punctuation from words, and putting the text into word-per-line format. Note that the verse code remains intact (2).

(2)
[MAT28_18/]
*yesu
akaja
kwao
,
akasema
nao
,
akawaambia
,
*nimepewa
mamlaka
yote
mbinguni
na

duniani
.$
[MAT28_19/]
*basi
,
enendeni
,
mkawafanye
mataifa
yote
kuwa
wanafunzi
,
mkiwabatiza
kwa
jina
la
*baba
,
na
la
*mwana
,
na
la
*roho_*mtakatifu
;
[MAT28_20/]
na
kuwafundisha
kuyashika
yote
niliyowaamuru
ninyi
;
na
tazama
,
mimi
nipo
pamoja
nanyi
siku
zote
,
hata

ukamilifu
wa
dahari
.$

6.3 Morphological analysis

The morphological analysis is carried out using either of the two methods, (a) the Finite State methods using Two-Level description (Koskenniemi 1983; Hurskainen 1992) or (b) the Two-Phase method using regular expressions (Hurskainen 2009). Both give precisely the same output. Note that the description includes also glosses in the target language, which is English in this case. Most of the words are ambiguous, and the inclusion of English glosses also increases ambiguity (3).

```
(3)
"<[MAT28_18/]>" LINE-CODE[3]
"<*yesu>"
       "*yesu" N PROPNAME AN HUM { *jesus } MALE
"<akaja>"
       "ja" V 1-SG3-SP VFIN { he } NARR:ka z [ja] { come } SV MONOSLB
       "ja" V 1-SG3-SP VFIN { she } NARR:ka z [ja] { come } SV MONOSLB
"<kwao>"
       "ao" PRON POSS 15-SG PL3 { their }
       "ao" PRON POSS 15-SG PL3 <PRE { them }
       "ao" PRON POSS 17-SG PL3 { for/at them }
       "ao" PRON POSS 17-SG PL3 { to them }
       "ao" PRON POSS 17-SG PL3 { from them }
       "ao" PRON POSS 17-SG PL3 { their } <LOC
"<,>"
       "," COMMA { , }
"<akasema>"
       "sema" V 1-SG3-SP VFIN { he } NARR:ka z [sema] { say } SVO
       "sema" V 1-SG3-SP VFIN { he } NARR:ka z [sema] { speak } SVO
       "sema" V 1-SG3-SP VFIN { he } NARR:ka z [sema] { scold } SVO
       "sema" V 1-SG3-SP VFIN { she } NARR:ka z [sema] { say } SVO
       "sema" V 1-SG3-SP VFIN { she } NARR:ka z [sema] { speak } SVO
       "sema" V 1-SG3-SP VFIN { she } NARR:ka z [sema] { scold } SVO
"<nao>"
       "nao" CC PRON PERS CC-PL 2-PL { and they }
       "nao" CC PRON PERS CC-PL 2-PL { also }
       "nao" CC PRON PERS CC-PL 2-PL { with them }
       "nao" CC PRON PERS CC-PL NO-PRON-GLOSS 2-PL { and }
       "nao" CC PRON CC-SG 3/4-SG { and it }
       "nao" CC PRON CC-SG 3/4-SG { also }
       "nao" CC PRON CC-SG 3/4-SG { with it }
       "nao" CC PRON CC-SG NO-PRON-GLOSS 3/4-SG { and }
       "nao" CC PRON CC-SG 11-SG { and it }
       "nao" CC PRON CC-SG 11-SG { also }
       "nao" CC PRON CC-SG 11-SG { with it }
```

[3] A list of tags appearing in the examples is in Appendix I.

```
       "nao" CC PRON CC-SG NO-PRON-GLOSS 11-SG { and }
       "nao" CC PRON CC-PL 2-PL { by them }
       "nao" CC PRON CC-SG 3/4-SG { by it }
       "nao" CC PRON CC-SG 11-SG { by it }
"<,>"
       "," COMMA { , }
"<akawaambia>"
       "ambia" V 1-SG3-SP VFIN { he } NARR:ka 2-PL2-OBJ OBJ { you } z
[amba] { tell } PREFR SVOO
       "ambia" V 1-SG3-SP VFIN { he } NARR:ka 2-PL2-OBJ OBJ { you } z
[amba] { say } SVO EXT: APPL :EXT
       "ambia" V 1-SG3-SP VFIN { he } NARR:ka 2-PL3-OBJ OBJ { them } z
[amba] { tell } PREFR SVOO
       "ambia" V 1-SG3-SP VFIN { he } NARR:ka 2-PL3-OBJ OBJ { them } z
[amba] { say } SVO EXT: APPL :EXT
       "ambia" V 1-SG3-SP VFIN { she } NARR:ka 2-PL2-OBJ OBJ { you } z
[amba] { tell } PREFR SVOO
       "ambia" V 1-SG3-SP VFIN { she } NARR:ka 2-PL2-OBJ OBJ { you } z
[amba] { say } SVO EXT: APPL :EXT
       "ambia" V 1-SG3-SP VFIN { she } NARR:ka 2-PL3-OBJ OBJ { them } z
[amba] { tell } PREFR SVOO
       "ambia" V 1-SG3-SP VFIN { she } NARR:ka 2-PL3-OBJ OBJ { them } z
[amba] { say } SVO EXT: APPL :EXT
"<,>"
       "," COMMA { , }
"<*nimepewa>"
       "pewa" V CAP 1-SG1-SP VFIN { *i } PERF:me z [pa] { give } PREFR
SVO PASS
       "pewa" V CAP 1-SG1-SP VFIN { *i } PERF:me z [pea] { mature } SV
EXT: PASS :EXT
       "pewa" V CAP 1-SG1-SP VFIN { *i } PERF:me z [pa] { give } SVOO
MONOSLB EXT: PASS :EXT
"<mamlaka>"
       "mamlaka" N 6-PLSG { the } { authority } AR
       "mamlaka" N 9/10-SG { the } { authority } AR
       "mamlaka" N 9/10-PL { the } { authority } AR
"<yote>"
       "ote" PRON :OTE 3/4-PL { all }
       "ote" PRON :OTE 5/6-PL { all }
       "ote" PRON :OTE 6-PLSG { all }
       "ote" PRON :OTE 9/10-SG { all }
"<mbinguni>"
       "mbingu" N 9/10-SG { the } { heavens } LOC { in }
       "mbingu" N 9/10-SG { the } { heavens } LOC { on }
       "mbingu" N 9/10-PL { the } { heavens } LOC { in }
       "mbingu" N 9/10-PL { the } { heavens } LOC { on }
"<na>"
       "na" CC { and }
       "na" AG-PART { by }
       "na" PREP { with }
       "na" NA-POSS { of }
       "na" ADV NOART { past }
"<duniani>"
       "dunia" N 9/10-SG { the } { :earth } PLACE AR LOC { in }
       "dunia" N 9/10-SG { the } { :earth } PLACE AR LOC { on }
       "dunia" N 9/10-SG { the } { :world } PLACE AR LOC { in }
       "dunia" N 9/10-SG { the } { :world } PLACE AR LOC { on }
```

```
       "dunia" N 9/10-PL { the } { :earth } PLACE AR LOC { in }
       "dunia" N 9/10-PL { the } { :earth } PLACE AR LOC { on }
       "dunia" N 9/10-PL { the } { :world } PLACE AR LOC { in }
       "dunia" N 9/10-PL { the } { :world } PLACE AR LOC { on }
"<.$>"
       ".$" { .$ } **CLB
```

6.4 Morphological disambiguation and syntactic mapping

Disambiguation is performed using a Constraint Grammar parser (Karlsson 1995; Tapanainen 1996, 1999; Hurskainen 1996, 2004a), which also performs syntactic mapping. Only one reading of each token is left, and the syntax is described as a surface syntax, without constructing full dependency trees. The article in English is defectively described; the default is a definite article. Fortunately this problem disappears in further phases of translation, because no African language has an article. An example of a disambiguated verse is in (4).

(4)

```
"<[MAT28_18/]>" LINE-CODE
"<*yesu>"
       "*yesu" N PROPNAME AN HUM { *jesus } MALE @SUBJ
"<akaja>"
       "ja" V 1-SG3-SP VFIN NO-SP-GLOSS NARR:ka z [ja] { come } SV
MONOSLB @FMAINVintr
"<kwao>"
       "ao" PRON POSS 17-SG PL3 { to them } @PRON-COMPL
"<,>"
       "," COMMA { , }
"<akasema>"
       "sema" V 1-SG3-SP VFIN { he } NARR:ka z [sema] { speak } SVO
@FMAINVtr+OBJ>
"<nao>"
       "nao" CC PRON PERS CC-PL 2-PL { with them } @OBJ
"<,>"
       "," COMMA { , }
"<akawaambia>"
       "ambia" V 1-SG3-SP VFIN { he } NARR:ka 2-PL3-OBJ OBJ { them } z
[amba] { tell } PREFR SVOO @FMAINVintr
"<,>"
       "," COMMA { , }
"<*nimepewa>"
       "pewa" V 1-SG1-SP VFIN { *i } PERF:me z [pa] { give } PREFR SVO
PASS CAP @FMAINVtr+OBJ>
"<mamlaka>"
       "mamlaka" N 6-PLSG { the } { authority } AR @OBJ
"<yote>"
       "ote" PRON :OTE 5/6-PL { all } @<PRON
"<mbinguni>"
       "mbingu" N 9/10-SG { the } { heavens } LOC { in } @NLOC
"<na>"
       "na" CC { and } @CC
"<duniani>"
       "dunia" N 9/10-SG { the } { :earth } PLACE AR LOC { on } @NLOC
```

```
"<.$>"
       ".$" { .$ } **CLB
```

6.5 Isolation of multi-word expressions

The concept 'multi-word expression' (MWE) is very difficult to define, because its definition depends on the type of two languages compared. What is a MWE in regard to English, is often not a MWE in regard to Luganda, because Bantu languages have much in common in formulating concepts. Therefore, the component of constructing MWEs in translating between Swahili and English does not apply as such in other translation tasks.

Also what is a single-word expression in Swahili is often a MWE in English. Typical and very frequent cases are inflected verbs, where Swahili uses prefixes and suffixes in expressing various functions and English uses separate words.

Therefore, it seems that the task of isolating MWEs in translating between Bantu languages is minimal compared with translation between structurally very different languages.

6.6 Removing glosses other than stems

When translating into Luganda, English glosses for morphemes, except for stems, can be removed, because the grammatical information is present already in the tag[4]. A slightly modified version of the disambiguation result (4) is in (5), where unnecessary glosses have been removed, and the reading of each word is on one line, surrounded by parentheses.

```
(5)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME AN HUM { *jesus } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP VFIN NO-SP-GLOSS NARR:ka z [ja] { come } SV
MONOSLB @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { to them } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP VFIN NARR:ka z [sema] { speak } SVO
@FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { with them } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP VFIN NARR:ka 2-PL3-OBJ OBJ z [amba] {
tell } PREFR SVOO @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP VFIN PERF:me z [pa] { get } PREFR SVO
PASS CAP @FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { authority } AR @OBJ )
```

[4] In fact, the English glosses of grammatical features could be excluded already in constructing the morphological lexicon. However, the process of removing them afterwards is demonstrated here for showing that (a) it is easy to do, and that (b) it is important to maintain a mother lexicon, which contains all notations possibly needed in any application. Removing information is easy; inserting information is a tough job.

```
( "<yote>" "ote" PRON :OTE 5/6-PL { all } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { heavens } LOC @NLOC )
( "<na>" "na" CC { and } @CC )
( "<duniani>" "dunia" N 9/10-SG { :earth } PLACE AR LOC @NLOC )
( "<.$>" ".$" { . } **CLB )
```

The output is further simplified by removing such tags that are not needed in further processing into Luganda (6).

```
(6)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { *jesus } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP NARR:ka { come } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { to them } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP NARR:ka { speak } @FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { with them } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP NARR:ka 2-PL3-OBJ { tell }
@FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP PERF:me { get } PASS CAP @FMAINVtr+OBJ>
)
( "<mamlaka>" "mamlaka" N 6-PLSG { authority } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { all } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { heavens } LOC @NLOC )
( "<na>" "na" CC { and } @CC )
( "<duniani>" "dunia" N 9/10-SG { :earth } PLACE LOC @NLOC )
( "<.$>" ".$" { . } **CLB )
```

6.7 Lexical transfer

Assuming that there is a transfer lexicon available with Luganda equivalents for English, the English stem glosses are compensated with stem glosses in Luganda, as in (7). Note that nouns have also a tag for noun group affiliation. This is not necessarily the same as in the source language. For this reason, the noun group code of the target language is inserted after the stem gloss. A colon ':' In front of stems is a temporary diacritic for indicating the beginning of the stem.

```
(7)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :*yesu [1/--SG] } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP NARR:ka { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP NARR:ka { :gambA } @FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { :nabo } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP NARR:ka 2-PL3-OBJ { :gambA }
@FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP PERF:me { :peeredwA } PASS CAP
@FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { :yinza [-/14] } @OBJ )
```

```
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { :ggulu [9/10] } LOC @NLOC )
( "<na>" "na" CC { :nE } @CC )
( "<duniani>" "dunia" N 9/10-SG { :nsi [9/10] } PLACE LOC @NLOC )
( "<.$>" ".$" { . } **CLB )
```

6.8 Handling divergent noun class affiliations

Cases, where the noun class affiliation of the SL and the TL may be different, are potentially all nouns, subject prefixes and object prefixes of the verb, adjectives, and genitive constructions. In practice, the system is constructed so that for each class marker inherited from the SL in a corresponding class marker in the TL. The processes are described below.

6.8.1 Nouns

The tags for noun group affiliation (after the stem gloss) are converted to precise tags indicating whether the noun is in singular or in plural. This is done on the basis of the information inherited from the source language (8). If the noun is a proper name, this phase is not needed, because the singular/plural distinction does not apply.

```
(8)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :*yesu [1/--SG] } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP NARR:ka { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP NARR:ka { :gambA } @FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { :nabo } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP NARR:ka 2-PL3-OBJ { :gambA }
@FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP PERF:me { :peeredwA } PASS CAP
@FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { :yinza [-/14-PL] } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { :ggulu [9/10-SG] } LOC @NLOC )
( "<na>" "na" CC { :nE } @CC )
( "<duniani>" "dunia" N 9/10-SG { :nsi [9/10-SG] } PLACE LOC @NLOC )
( "<.$>" ".$" { . } **CLB )
```

6.8.2 Verbs

The conversion of noun tags is quite simple as shown in (7) and (8). The conversion of the subject prefix and object prefix of the verb is more complicated, because their form depends on the subject and object respectively. Therefore, we need a wider context for deciding the noun class of the subject prefix and object prefix of the verb.

The problem can be solved in at least two ways. In one solution, the noun class of the subject prefix, inherited from Swahili, is replaced by the noun class found in the

subject of the target language in the clause. Finding the correct noun class for the object prefix can be done by looking for the noun class of the object in target language. Often the object prefix is obvious, however. Rules for managing subject prefixes and object prefixes in the target language can be implemented using regular expressions and substitution.

Another solution, perhaps easier to implement, is to write rules in a CG parser. Here I will demonstrate how this solution was implemented. Because the primary operators in the CG parser function on the principle of selection and removal, it is useful first to add all subject prefix candidates to the reading. An example of this is in (9)[5].

```
(9)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :*yesu [1/--SG] } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP
        [1-SG1-SP]
        [1-SG2-SP]
        [1-SG3-SP]
        [2-PL1-SP]
        [2-PL2-SP]
        [2-PL3-SP]
        [3-SG-SP]
        [4-PL-SP]
        [5-SG-SP]
        [6-PL-SP]
        [7-SG-SP]
        [8-PL-SP]
        [9-SG-SP]
        [10-PL-SP]
        [11-SG-SP]
        [12-SG-SP]
        [13-PL-SP]
        [14-PL-SP]
        [15-SG-SP]
        [16-SG-SP]
        [17-SG-SP]
        [18-SG-SP]
        [19-SG-SP]
        [20-SG-SP]
        [22-PL-SP]
        [23-SG-SP]
"<x>" "x" NARR:ka { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
```

All candidates of the object prefix are also added (10).

```
(10)
( "<akawaambia>" "ambia" V 1-SG3-SP
        [1-SG1-SP]
```

[5] The CG parser expects that each line in the cohort is preceded by a tab. The reading is formatted to meet this requirement. Also temporary tags, such as "<x>" and "x", are added so that the parser functions correctly.

```
        [1-SG2-SP]
        [1-SG3-SP]
        [2-PL1-SP]
        [2-PL2-SP]
        [2-PL3-SP]
        [3-SG-SP]
        [4-PL-SP]
        [5-SG-SP]
        [6-PL-SP]
        [7-SG-SP]
        [8-PL-SP]
        [9-SG-SP]
        [10-PL-SP]
        [11-SG-SP]
        [12-SG-SP]
        [13-PL-SP]
        [14-PL-SP]
        [15-SG-SP]
        [16-SG-SP]
        [17-SG-SP]
        [18-SG-SP]
        [19-SG-SP]
        [20-SG-SP]
        [22-PL-SP]
        [23-SG-SP]
"<x>" "x" NARR:ka 2-PL3-OBJ
        [1-SG1-OBJ]
        [1-SG2-OBJ]
        [1-SG3-OBJ]
        [2-PL1-OBJ]
        [2-PL2-OBJ]
        [2-PL3-OBJ]
        [3-SG-OBJ]
        [4-PL-OBJ]
        [5-SG-OBJ]
        [6-PL-OBJ]
        [7-SG-OBJ]
        [8-PL-OBJ]
        [9-SG-OBJ]
        [10-PL-OBJ]
        [11-SG-OBJ]
        [12-SG-OBJ]
        [13-PL-OBJ]
        [14-PL-OBJ]
        [15-SG-OBJ]
        [16-SG-OBJ]
        [17-SG-OBJ]
        [18-SG-OBJ]
        [19-SG-OBJ]
        [20-SG-OBJ]
        [22-PL-OBJ]
        [23-SG-OBJ]
"<x>" "x" { :gambA } @FMAINVintr )
( "<,>" "," COMMA { , } )
```

The infinitive form of the verb can also have an object prefix, and also this should be handled. This is done in the same way as the object prefix of the finite verbs. An example is in (11), where there are two infinitive verbs, one subordinated to the other. In the example, there is also a finite verb with object prefix and overt object.

```
(11)
( "<[MAT28_20/]>" LINE-CODE )
( "<na>" "na" CC { :nE } @CC )
( "<kuwafundisha>" "fundisha" V INF 2-PL3-OBJ
       [1-SG1-OBJ]
       [1-SG2-OBJ]
       [1-SG3-OBJ]
       [2-PL1-OBJ]
       [2-PL2-OBJ]
       [2-PL3-OBJ]
       [3-SG-OBJ]
       [4-PL-OBJ]
       [5-SG-OBJ]
       [6-PL-OBJ]
       [7-SG-OBJ]
       [8-PL-OBJ]
       [9-SG-OBJ]
       [10-PL-OBJ]
       [11-SG-OBJ]
       [12-SG-OBJ]
       [13-PL-OBJ]
       [14-PL-OBJ]
       [15-SG-OBJ]
       [16-SG-OBJ]
       [17-SG-OBJ]
       [18-SG-OBJ]
       [19-SG-OBJ]
       [20-SG-OBJ]
       [22-PL-OBJ]
       [23-SG-OBJ]
"<x>" "x" { :yigilizA } CAUS @-FMAINV-n )
( "<kuyashika>" "shika" V INF 6-PL-OBJ
       [1-SG1-OBJ]
       [1-SG2-OBJ]
       [1-SG3-OBJ]
       [2-PL1-OBJ]
       [2-PL2-OBJ]
       [2-PL3-OBJ]
       [3-SG-OBJ]
       [4-PL-OBJ]
       [5-SG-OBJ]
       [6-PL-OBJ]
       [7-SG-OBJ]
       [8-PL-OBJ]
       [9-SG-OBJ]
       [10-PL-OBJ]
       [11-SG-OBJ]
       [12-SG-OBJ]
       [13-PL-OBJ]
       [14-PL-OBJ]
       [15-SG-OBJ]
```

```
        [16-SG-OBJ]
        [17-SG-OBJ]
        [18-SG-OBJ]
        [19-SG-OBJ]
        [20-SG-OBJ]
        [22-PL-OBJ]
        [23-SG-OBJ]
"<x>" "x" { :kwatA } @-FMAINV-n )
( "<yote>" "ote" PRON :OTE 6-PL { :onna } @OBJ )
( "<niliyowaamuru>" "amuru" V 1-SG1-SP
        [1-SG1-SP]
        [1-SG2-SP]
        [1-SG3-SP]
        [2-PL1-SP]
        [2-PL2-SP]
        [2-PL3-SP]
        [3-SG-SP]
        [4-PL-SP]
        [5-SG-SP]
        [6-PL-SP]
        [7-SG-SP]
        [8-PL-SP]
        [9-SG-SP]
        [10-PL-SP]
        [11-SG-SP]
        [12-SG-SP]
        [13-PL-SP]
        [14-PL-SP]
        [15-SG-SP]
        [16-SG-SP]
        [17-SG-SP]
        [18-SG-SP]
        [19-SG-SP]
        [20-SG-SP]
        [22-PL-SP]
        [23-SG-SP]
"<x>" "x" PAST 3/4-PL-REL 2-PL2-OBJ
        [1-SG1-OBJ]
        [1-SG2-OBJ]
        [1-SG3-OBJ]
        [2-PL1-OBJ]
        [2-PL2-OBJ]
        [2-PL3-OBJ]
        [3-SG-OBJ]
        [4-PL-OBJ]
        [5-SG-OBJ]
        [6-PL-OBJ]
        [7-SG-OBJ]
        [8-PL-OBJ]
        [9-SG-OBJ]
        [10-PL-OBJ]
        [11-SG-OBJ]
        [12-SG-OBJ]
        [13-PL-OBJ]
        [14-PL-OBJ]
        [15-SG-OBJ]
        [16-SG-OBJ]
```

```
       [17-SG-OBJ]
       [18-SG-OBJ]
       [19-SG-OBJ]
       [20-SG-OBJ]
       [22-PL-OBJ]
       [23-SG-OBJ]
"<x>" "x" { :lagilA } @FMAINVtr+OBJ> )
( "<ninyi>" "ninyi" PRON PERS PL2 { :mmwe } @OBJ )
( "<;>" ";" SEMI-COLON **CLB )
```

6.8.3 Genitive constructions

Genitive constructions have a genitive connector, where the form of the connector depends on the class of the preceding noun. Also this is handled by introducing first all connector candidates. This is done in a different way than in introducing subject and object prefix tags. The tags in (12) combine the surface form and the class affiliation[6].

```
(12)
( "<ukamilifu>" "ukamilifu" N 11-SG { perfection } )
( "<wa>" "wa" GEN-CON 11-SG
       wa-1
       ba-2
       gwa-3
       gya-4
       lya-5
       ga-6
       kya-7
       bya-8
       ya-9
       za-10
       lwa-11
       ka-12
       bwa-14
       gwa
       kwa-15
       ga-23
"<x>" "x" @GCON )
( "<dahari>" "dahari" N 9/10-SG { aeon } @<GN )
```

6.8.4 Disambiguation of tags

When all the needed tags are added, disambiguation rules written in CG are applied. The rules are ordered so that first an overt subject is looked for in the clause. If it is found, the noun class of that subject is selected among the subject prefix candidates of the verb. If no subject is found, the original subject prefix tag inherited from the source language is selected.

[6] The latter is needed for disambiguation.

The same procedure is applied to choosing the object prefix. The class of the overt object is selected as the class of the object prefix. If no object is found, the object prefix tag of the source language is selected.

The class of the genitive connector is selected according to the preceding noun. Example (13) illustrates the disambiguation of all the cases in (9 - 12). The disambiguated instances can be seen as non-continuous lines.

```
(13)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :*yesu [1/--SG] } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP
       [1-SG3-SP]
"<x>" "x" NARR:ka { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP
       [1-SG3-SP]
"<x>" "x" NARR:ka { :gambA } @FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { :nabo } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP
       [1-SG3-SP]
"<x>" "x" NARR:ka 2-PL3-OBJ
       [2-PL3-OBJ]
"<x>" "x" { :gambA } @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP
       [1-SG1-SP]
"<x>" "x" PERF:me { :peeredwA } PASS CAP @FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { :yinza [-/14] } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { :ggulu [9/10-SG] } LOC @NLOC )
( "<na>" "na" CC { :nE } @CC )
( "<duniani>" "dunia" N 9/10-SG { :nsi [9/10-SG] } PLACE LOC @NLOC )
( "<.$>" ".$" { . } **CLB )
 ( "<[MAT28_19/]>" LINE-CODE )
( "<*basi>" "basi" ADV { :kale } CAP @ADVL )
( "<,>" "," COMMA { , } )
( "<enendeni>" "enenda" V IMP { :gendA } IMP-PL2 @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<mkawafanye>" "fanya" V 2-PL2-SP
       [2-PL2-SP]
"<x>" "x" CONSEQ:ka-e 2-PL3-OBJ
       [2-PL3-OBJ]
"<x>" "x" { :fuulA } @FMAINVtr+OBJ> )
( "<mataifa>" "taifa" N 5/6-PL { :wanga [-/14] } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<kuwa>" "wa" V INF MOD-CAN { :li } AUX-WA @-FMAINV-n )
( "<wanafunzi>" "mwanafunzi" N 2-PL { :yigilizwa [2-PL] } @NCOMPL )
( "<,>" "," COMMA { , } )
( "<mkiwabatiza>" "batiza" V 2-PL2-SP
       [2-PL2-SP]
"<x>" "x" COND-IF 2-PL3-OBJ
       [2-PL3-OBJ]
"<x>" "x" { :batizA } @FMAINVtr-OBJ> )
```

```
( "<kwa>" "kwa" PREP { mu:- } @PREP )
( "<jina>" "jina" N 5/6-SG { :linnya [5/6-SG] } @NCOMPL )
( "<la>" "la" GEN-CON 5-SG
      lya-5
"<x>" "x" @GCON )
( "<*baba>" "baba" N TITLE { :*kitaffe [1/--SG] } @<GN )
( "<,>" "," COMMA { , } )
( "<na>" "na" CC { :nE } @CC )
( "<la>" "la" GEN-CON 5-SG
      lya-5
"<x>" "x" )
( "<*mwana>" "mwana" N PROPNAME { :*omwana [1/--SG] } MALE CAP @<GN )
( "<,>" "," COMMA { , } )
( "<na>" "na" CC { :nE } @CC )
( "<la>" "la" GEN-CON 5-SG
      lya-5
"<x>" "x" )
( "<*roho_*mtakatifu>" "roho_*mtakatifu" N PROPNAME { :*omwoyo_*omutukuvu
[1/--SG] } MALE CAP @<GN )
( "<;>" ";" SEMI-COLON **CLB )
 ( "<[MAT28_20/]>" LINE-CODE )
( "<na>" "na" CC { :nE } @CC )
( "<kuwafundisha>" "fundisha" V INF 2-PL3-OBJ
      [2-PL3-OBJ]
"<x>" "x" { :yigilizA } CAUS @-FMAINV-n )
( "<kuyashika>" "shika" V INF 6-PL-OBJ
      [6-PL-OBJ]
"<x>" "x" { :kwatA } @-FMAINV-n )
( "<yote>" "ote" PRON :OTE 6-PL { :onna } @OBJ )
( "<niliyowaamuru>" "amuru" V 1-SG1-SP
      [1-SG1-SP]
"<x>" "x" PAST 3/4-PL-REL 2-PL2-OBJ
      [2-PL2-OBJ]
"<x>" "x" { :lagilA } @FMAINVtr+OBJ> )
( "<ninyi>" "ninyi" PRON PERS PL2 { :mmwe } @OBJ )
( "<;>" ";" SEMI-COLON **CLB )
( "<na>" "na" CC { :nE } @CC )
( "<tazama>" "tazama" V IMP { :labA } @FMAINVtr-OBJ> )
( "<,>" "," COMMA { , } )
( "<mimi>" "mimi" PRON PERS SG1 { :nze } @SUBJ )
( "<nipo>" "nipo" V-BE SG1-SP { :ndi } LOC-16 @FMAINVintr )
( "<pamoja>" "pamoja" ADV { :wamu } @ADVL )
( "<nanyi>" "nanyi" PRON PERS CC-PL 2-PL2 { :nammwe } @PRON-COMPL )
( "<siku>" "siku" N 9/10-PL { :ennaku [9/10-PL] } TIME @NCOMPL )
( "<zote>" "ote" PRON :OTE 9/10-PL { :onna } @<PRON )
( "<,>" "," COMMA { , } )
( "<hata>" "hata" ADV { until } @ADVL )
( "<ukamilifu>" "ukamilifu" N 11-SG { perfection } @NCOMPL )
( "<wa>" "wa" GEN-CON 11-SG
      lwa-11
"<x>" "x" @GCON )
( "<dahari>" "dahari" N 9/10-SG { aeon } @<GN )
( "<.$>" ".$" { . } **CLB )
```

After some formatting, each word-form is again on its own line, surrounded by brackets (14). Note that the tags of the SL and of the TL are still there for making the final choice possible.

```
(14)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :*yesu [1/--SG] } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP [1-SG3-SP] NARR:ka { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP [1-SG3-SP] NARR:ka { :gambA }
@FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { :nabo } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP [1-SG3-SP] NARR:ka 2-PL3-OBJ [2-PL3-
OBJ] { :gambA } @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP [1-SG1-SP] PERF:me { :peeredwA } PASS
CAP @FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { :yinza [-/14] } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { :ggulu [9/10-SG] } LOC @NLOC )
( "<na>" "na" CC { :nE } @CC )
( "<duniani>" "dunia" N 9/10-SG { :nsi [9/10-SG] } PLACE LOC @NLOC )
( "<.$>" ".$" { . } **CLB )
( "<[MAT28_19/]>" LINE-CODE )
( "<*basi>" "basi" ADV { :kale } CAP @ADVL )
( "<,>" "," COMMA { , } )
( "<enendeni>" "enenda" V IMP { :gendA } IMP-PL2 @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<mkawafanye>" "fanya" V 2-PL2-SP [2-PL2-SP] CONSEQ:ka-e 2-PL3-OBJ [2-
PL3-OBJ] { :fuulA } @FMAINVtr+OBJ> )
( "<mataifa>" "taifa" N 5/6-PL { :wanga [-/14] } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<kuwa>" "wa" V INF MOD-CAN { :li } AUX-WA @-FMAINV-n )
( "<wanafunzi>" "mwanafunzi" N 2-PL { :yigilizwa [1/2-PL] } @NCOMPL )
( "<,>" "," COMMA { , } )
( "<mkiwabatiza>" "batiza" V 2-PL2-SP [2-PL2-SP] COND-IF 2-PL3-OBJ [2-
PL3-OBJ] { :batizA } @FMAINVtr-OBJ> )
( "<kwa>" "kwa" PREP { mu:- } @PREP )
( "<jina>" "jina" N 5/6-SG { :linnya [5/6-SG] } @NCOMPL )
( "<la>" "la" GEN-CON 5/6-SG lya-5 @GCON )
( "<*baba>" "baba" N TITLE { :*kitaffe [1/--SG] } @<GN )
( "<,>" "," COMMA { , } )
( "<na>" "na" CC { :nE } @CC )
( "<la>" "la" GEN-CON 5/6-SG lya-5 @GCON )
( "<*mwana>" "mwana" N PROPNAME { :*omwana [1/--SG] } MALE CAP @<GN )
( "<,>" "," COMMA { , } )
( "<na>" "na" CC { :nE } @CC )
( "<la>" "la" GEN-CON 5/6-SG lya-5 @GCON )
( "<*roho_*mtakatifu>" "roho_*mtakatifu" N PROPNAME { :*omwoyo_*omutukuvu
[1/--SG] } MALE CAP @<GN )
( "<;>" ";" SEMI-COLON **CLB )
( "<[MAT28_20/]>" LINE-CODE )
( "<na>" "na" CC { :nE } @CC )
```

```
( "<kuwafundisha>" "fundisha" V INF 2-PL3-OBJ [2-PL3-OBJ] { :yigilizA }
CAUS @-FMAINV-n )
( "<kuyashika>" "shika" V INF 5/6-PL-OBJ [6-PL-OBJ] { :kwatA } @-FMAINV-n
)
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @OBJ )
( "<niliyowaamuru>" "amuru" V 1-SG1-SP [1-SG1-SP] PAST 3/4-PL-REL 2-PL2-
OBJ [2-PL2-OBJ] { :lagilA } @FMAINVtr+OBJ> )
( "<ninyi>" "ninyi" PRON PERS PL2 { :mmwe } @OBJ )
( "<;>" ";" SEMI-COLON **CLB )
( "<na>" "na" CC { :nE } @CC )
( "<tazama>" "tazama" V IMP { :labA } @FMAINVtr-OBJ> )
( "<,>" "," COMMA { , } )
( "<mimi>" "mimi" PRON PERS SG1 { :nze } @SUBJ )
( "<nipo>" "nipo" V-BE SG1-SP { :ndi } LOC-16 @FMAINVintr )
( "<pamoja>" "pamoja" ADV { :wamu } @ADVL )
( "<nanyi>" "nanyi" PRON PERS CC-PL 2-PL2 { :nammwe } @PRON-COMPL )
( "<siku>" "siku" N 9/10-PL { :ennaku [9/10-PL] } TIME @NCOMPL )
( "<zote>" "ote" PRON :OTE 9/10-PL { :onna } @<PRON )
( "<,>" "," COMMA { , } )
( "<hata>" "hata" ADV { until } @ADVL )
( "<ukamilifu>" "ukamilifu" N 11-SG { perfection } @NCOMPL )
( "<wa>" "wa" GEN-CON 11-SG lwa-11 @GCON )
( "<dahari>" "dahari" N 9/10-SG { aeon } @<GN )
( "<.$>" ".$" { . } **CLB )
```

6.9 Handling divergent verb structures

In some verb-forms, such as subjunctive, imperative, and consecutive, the final *a* changes into *e* (15).

```
(15)
( "<[MAT28_19/]>" LINE-CODE )
( "<*basi>" "basi" ADV { :kale } CAP @ADVL )
( "<,>" "," COMMA { , } )
( "<enendeni>" "enenda" V IMP { :gende } IMP-PL2 @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<mkawafanye>" "fanya" V 2-PL2-SP [2-PL2-SP] CONSEQ:ka-e 2-PL3-OBJ [2-
PL3-OBJ] { :fuule } @FMAINVtr+OBJ> )
( "<mataifa>" "taifa" N 5/6-PL { :wanga [-/14] } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<kuwa>" "wa" V INF MOD-CAN { :li } AUX-WA @-FMAINV-n )
( "<wanafunzi>" "mwanafunzi" N 1/2-PL { :yigilizwa [1/2-PL] } @NCOMPL )
```

6.10 Converting noun tags to lexical form

In the next step, noun tags of the target language are converted into lexical form[7] and moved in front of the noun stem (16).

```
(16)
```

[7] Note that these tags are not yet necessarily surface forms, because morpho-phonological rules have not yet been applied.

```
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :0-:*yesu } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP [1-SG3-SP] NARR:ka { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP [1-SG3-SP] NARR:ka { :gambA }
@FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { :nabo } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP [1-SG3-SP] NARR:ka 2-PL3-OBJ [2-PL3-
OBJ] { :gambA } @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP [1-SG1-SP] PERF:me { :peeredwA } PASS
CAP @FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { :AmA-:yinza } @OBJ )
( "<yote>" "ote" PRON :OTE 5/6-PL { :onna } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { :E0-:ggulu } LOC @NLOC )
( "<na>" "na" CC { :nE } @CC )
( "<duniani>" "dunia" N 9/10-SG { :E0-:nsi } PLACE LOC @NLOC )
```

6.11 Converting verb tags to lexical form

So far the prefixes of the verbs have been described by means of tags. The tags for the target language are within square brackets, and the tags for the source language are without brackets. In the next step the verb tags of the target language are converted into lexical forms, and the square brackets are converted to curly brackets to indicate a lexical gloss (17).

```
(17)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :0-:*yesu } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP { a- } { A- } { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP { a- } { A- } { :gambA } @FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { :nabo } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP { a- } { A- } 2-PL3-OBJ { bA- } {
:gambA } @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP { n- } { 0- } { :peeredwA } PASS CAP
@FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { :AmA-:yinza } @OBJ )
( "<yote>" "ote" PRON { gI- } { :onna } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { :E0-:ggulu } LOC @NLOC )
( "<na>" "na" CC { :nE } @CC )
( "<duniani>" "dunia" N 9/10-SG { :E0-:nsi } PLACE LOC @NLOC )
( "<.$>" ".$" { . } **CLB )
```

7 Re-ordering constituents

In the Swahili-to-English translation system, there is a large module for re-ordering constituents of the source language to meet the requirements of the target language. In

addition to word order, also word-internal constituents need re-ordering. This is perhaps the most difficult module of the whole translation system to implement in the Swahili-to-English translation system.

When we translate from Swahili to Luganda, the need for re-ordering constituents is minimal compared with Swahili-to-English translation. In the examples dealt with in this paper, no reordering rules are needed, except for the locative and some other grammatical structures, where the structure is realized in a different way in Luganda. In Swahili, the locative of nouns is marked with the suffix *-ni*, while Luganda uses a separate locative prefix of one of the three locative classes. Also the plural imperative (PL2) is marked in Swahili by the suffix *-ni*, but Luganda uses the prefix *mU-*.

These are examples of cases, where the order of constituents does not stay the same between two Bantu languages. These and other re-ordering cases are handled using a rule set designed for this purpose using regular expressions.

The locative and imperative forms of the target language are produced on the lexical level (18).

```
(18)
( "<[MAT28_18/]>" LINE-CODE )
( "<*yesu>" "yesu" N PROPNAME { :0-:*yesu } MALE CAP @SUBJ )
( "<akaja>" "ja" V 1-SG3-SP { a- } { A- } { :jja } @FMAINVintr )
( "<kwao>" "ao" PRON POSS 17-SG PL3 { :nabo } @PRON-COMPL )
( "<,>" "," COMMA { , } )
( "<akasema>" "sema" V 1-SG3-SP { a- } { A- } { :gambA } @FMAINVtr+OBJ> )
( "<nao>" "nao" PRON PERS CC-PL 2-PL { :nabo } @OBJ )
( "<,>" "," COMMA { , } )
( "<akawaambia>" "ambia" V 1-SG3-SP { a- } { A- } 2-PL3-OBJ { bA- } {
:gambA } @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<*nimepewa>" "pewa" V 1-SG1-SP { n- } { 0- } { :peeredwA } PASS CAP
@FMAINVtr+OBJ> )
( "<mamlaka>" "mamlaka" N 6-PLSG { :AmA-:yinza } @OBJ )
( "<yote>" "ote" PRON { gI- } { :onna } @<PRON )
( "<mbinguni>" "mbingu" N 9/10-SG { :ku/mu } { :E0-:ggulu } @NLOC )
( "<na>" "na" CC { :nE } @CC )
( "<duniani>" "dunia" N 9/10-SG { :ku/mu } { :E0-:nsi } @NLOC )
( "<.$>" ".$" { . } **CLB )
 ( "<[MAT28_19/]>" LINE-CODE )
( "<*basi>" "basi" ADV { :kale } CAP @ADVL )
( "<,>" "," COMMA { , } )
( "<enendeni>" "enenda" V IMP { mU- } { :gende } @FMAINVintr )
( "<,>" "," COMMA { , } )
( "<mkawafanye>" "fanya" V 2-PL2-SP { mU- } CONSEQ:ka-e 2-PL3-OBJ { bA- }
{ :fuule } @FMAINVtr+OBJ> )
( "<mataifa>" "taifa" N 5/6-PL { :AmA-:wanga } @OBJ )
( "<yote>" "ote" PRON { gI- } { :onna } @<PRON )
( "<kuwa>" "wa" V { OkU- } MOD-CAN { :li } AUX-WA @-FMAINV-n )
( "<wanafunzi>" "mwanafunzi" N 1/2-PL { :AbA-:yigilizwa } @NCOMPL )
```

8 From lexical to surface form

In this phase, all lexical morphemes of the target language have been constructed. Also they are in the order where they should appear. Some of the morphemes are already attached to the corresponding stems. Others are still as separate constituents, but they are located in the place where they should be in the final text. However, what is common to all morphemes is that they are still in lexical form subject to alternation rules.

8.1 Concatenating lexical morphemes into intermediate words

When all morphemes of the source language have been transferred into the forms of the target language, and the reordering of these lexical elements performed, grammatical tags can be removed. Everything else is removed except what is between curly brackets. As a result, the lexical morphemes are left, one morpheme per line, in the order where they should appear in surface text of the target language (19).

```
(19)
{ :0-:*yesu }
{ a- }
{ A- }
{ :jja }
{ :nabo }
{ , }
{ a- }
{ A- }
{ :gambA }
{ :nabo }
{ , }
{ a- }
{ A- }
{ bA- }
{ :gambA }
{ , }
{ n- }
{ 0- }
{ :peeredwA }
{ :AmA-:yinza }
{ gI- }
{ :onna }
{ :ku/mu }
{ :E0-:ggulu }
{ :nE }
{ :ku/mu }
{ :E0-:nsi }
{ . }
[MAT28_19/]
{ :*kale }
{ , }
{ mU- }
{ :gende }
{ , }
{ mU- }
{ bA- }
{ :fuule }
{ :AmA-:wanga }
{ gI- }
{ :onna }
{ OkU- }
{ :li }
{ :AbA-:yigilizwa }
{ , }
{ mU- }
{ bA- }
{ :batizA }
{ mu:- }
{ :E0D-:linnya }
{ :0-:*kitaffe }
{ , }
{ :nE }
{ :0-:*omwana }
{ , }
{ :nE }
{ :0-:*omwoyo_*omutukuvu }
[MAT28_20/]
{ :nE }
{ OkU- }
{ bA- }
{ :yigilizA }
{ OkU- }
{ gA- }
{ :kwatA }
{ gI- }
{ :onna }
{ n- }
{ A- }
{ bA- }
{ :lagilA }
```

```
{ :mmwe }                          { :E0-:ennaku }
{ :nE }                            { zI- }
{ :labe }                          { :onna }
{ , }                              { , }
{ :nze }                           { until }
{ :ndi }                           { perfection }
{ :wamu }                          { aeon }
{ :nammwe }                        { . }
```

As it is shown in (19), the morphemes of a verb are still on separate lines. These will be joined as a single string, as in (20).[8]

```
(20)
[MAT28_18/]
{ :0-:*yesu }
{ a-A-:jja }
{ :nabo }
{ , }
{ a-A-:gambA }
{ :nabo }
{ , }
{ a-A-bA-:gambA }
{ , }
{ n-0-:peeredwA }
{ :AmA-:yinza }
{ gI-:onna }
{ :ku/mu }
{ :E0-:ggulu }
{ :nE }
{ :ku/mu }
{ :E0-:nsi }
{ . }
[MAT28_19/]
{ :*kale }
{ , }
{ mU-:gende }
{ , }
{ mU-bA-:fuule }
{ :AmA-:wanga }
{ gI-:onna }
{ OkU-:li }
{ :AbA-:yigilizwa }
{ , }
{ mU-bA-:batizA }
{ :mu }
{ :linnya }
{ :0-:*kitaffe }
{ , }
{ :nE }
{ :0-:*omwana }
```

[8] A dash '-' attached to the morpheme indicates to which side the morpheme should be concatenated.

```
{ , }
{ :nE }
{ :0-:*omwoyo_*omutukuvu }
[MAT28_20/]
{ :nE }
{ OkU-bA-:yigilizA }
{ OkU-gA-:kwatA }
{ gI-:onna }
{ n-A-bA-:lagilA }
{ :mmwe }
{ :nE }
{ :labe }
{ , }
{ :nze }
{ :ndi }
{ :wamu }
{ :nammwe }
{ :E0-:ennaku }
{ zI-:onna }
{ , }
{ until }
{ perfection }
{ aeon }
{ . }
```

8.2 Applying morpho-phonological rules

Now the strings in the TL are in the format, where morpho-phonological alternation rules can be applied. After rule application, the strings start to resemble surface forms. They are not yet quite ready, as seen in (21)

(21)

```
[MAT28_18/]
{ ::*yesu }
{ a:jja }
{ :nabo }
{ , }
{ a:gamba }
{ :nabo }
{ , }
{ aba:gamba }
{ , }
{ m:peeredwa }
{ :ama:yinza }
{ g:onna }
{ :ku/mu }
{ :e:ggulu }
{ :ne }
{ :ku/mu }
{ :e:nsi }
{ . }
[MAT28_19/]
{ :*kale }
{ , }
{ mu:gende }
{ , }
{ muba:fuule }
{ :ama:wanga }
{ g:onna }
{ oku:li }
{ :aba:yigilizwa }
{ , }
{ muba:batiza }
{ :mu }
{ :linnya }
{ ::*kitaffe }
{ , }
{ :ne }
{ ::*omwana }
{ , }
{ :ne }
{ ::*omwoyo_*omutukuvu }
[MAT28_20/]
{ :ne }
{ okuba:yigiliza }
{ okuga:kwata }
```

```
{ g:onna }
{ naba:lagila }
{ :mmwe }
{ :ne }
{ :labe }
{ , }
{ :nze }
{ :ndi }
{ :wamu }
{ :nammwe }
{ e:nnaku }
{ z:onna }
{ , }
{ until }
{ perfection }
{ aeon }
{ . }
```

8.3 Marking un-translated words

The un-translated words are marked, after which the temporary colon is removed. Also stem-initial consonants are doubled, if they are not yet double consonants (22).

```
(22)
[MAT28_18/]
{ *yesu }
{ ajja }
{ nabo }
{ , }
{ agamba }
{ nabo }
{ , }
{ abagamba }
{ , }
{ mpeeredwa }
{ amayinza }
{ gonna }
{ ku/mu }
{ eggulu }
{ ne }
{ ku/mu }
{ ensi }
{ . }
[MAT28_19/]
{ *kale }
{ , }
{ mugende }
{ , }
{ mubafuule }
{ amawanga }
{ gonna }
{ okuli }
{ abayigilizwa }
{ , }
{ mubabatiza }
{ mu }
{ linnya }
{ *kitaffe }
{ , }
{ ne }
{ *omwana }
{ , }
{ ne }
{ *omwoyo_*omutukuvu }
[MAT28_20/]
{ ne }
{ okubayigiliza }
{ okugakwata }
{ gonna }
{ nabalagila }
{ mmwe }
{ ne }
{ labe }
{ , }
{ nze }
{ ndi }
{ wamu }
{ nammwe }
{ ennaku }
{ zonna }
{ , }
{ <until> }
{ <perfection> }
{ <aeon> }
{ . }
```

8.4 Final pruning

After some pruning and reformatting the final translated output string is produced as shown in (23).

(23)
[MAT28_18/] Yesu ajja nabo, agamba nabo, abagamba, mpeeredwa amayinza gonna ku/mu eggulu ne ku/mu ensi.
[MAT28_19/] Kale, mugende, mubafuule amawanga gonna okuli abayigilizwa, mubabatiza mu linnya Kitaffe, ne Omwana, ne Omwoyo Omutukuvu
[MAT28_20/] ne okubayigiliza okugakwata gonna nabalagila mmwe ne labe, nze ndi wamu nammwe ennaku zonna, <until> <perfection> <aeon>.

We see that the locative class in verse 18 is underspecified (ku/mu), because the ambiguity between classes 17 and 18 cannot be resolved. There are also three words remaining un-translated, *until*, *perfection*, and *aeon*, and they are clearly marked for helping in final editing. These words were left un-translated purposefully, for demonstrating how the system automatically marks un-translated items.

9 Comparison with other translations

When inspecting and editing the Luganda translation produced by the system, other translations can be made available next to each translated line. For example, the original Swahili text and its translation into English can be presented, as in (24).

(24)
Lug: [MAT28_18/] Yesu ajja nabo, agamba nabo, abagamba, mpeeredwa amayinza gonna ku/mu eggulu ne ku/mu ensi.
Swa: [MAT28_18/] Yesu akaja kwao, akasema nao, akawaambia, Nimepewa mamlaka yote mbinguni na duniani.
Eng: [MAT28_18/] Jesus came to them, he spoke with them, he told them, I have been given all authority in the heavens and on the earth.

Lug: [MAT28_19/] Kale, mugende, mubafuule amawanga gonna okuli abayigilizwa, mubabatiza mu linnya Kitaffe, ne Omwana, ne Omwoyo Omutukuvu
Swa: [MAT28_19/] Basi, enendeni, mkawafanye mataifa yote kuwa wanafunzi, mkiwabatiza kwa jina la Baba, na la Mwana, na la Roho Mtakatifu;
Eng: [MAT28_19/] then, go, and do all nations be the pupils, if/when you baptize them in the name of Father, and of Son, and of Holy Spirit;

Lug: [MAT28_20/] ne okubayigiliza okugakwata gonna nabalagila mmwe ne labe, nze ndi wamu nammwe ennaku zonna, <until> <perfection> <aeon>.
Swa: [MAT28_20/] na kuwafundisha kuyashika yote niliyowaamuru ninyi; na tazama, mimi nipo pamoja nanyi siku zote, hata ukamilifu wa dahari.
Eng: [MAT28_20/] and to teach them to take hold of all which I commanded you; and look, I am together with you all days, until the perfection of the aeon.

10 Discussion and conclusion

In this paper we have seen how SALAMA was implemented to the task of translating the Bible into another language. The system was implemented to Luganda as a target language. All major phases in the process were discussed and solutions to problems were demonstrated.

Because there was available material for constructing a transfer lexicon between English and Luganda, I used the English lexical glosses of SALAMA for replacing each gloss with a Luganda equivalent. If there would have been a transfer lexicon between Swahili and Luganda, the transfer could have been made directly from Swahili to Luganda. But because such a lexicon does not exist, the transfer had to be made via English.

The solution made here for implementing the lexical transfer has wider implications in other translation tasks. English functions in large areas of Africa as a lingua franca, and the existing lexicons are usually compiled between English and the local language concerned. Therefore, the problem of transfer language that occurred in this study is likely to occur in many other translation tasks. In practice, often the transfer has to be made from English to the target language.

In conclusion, the rule-based translation method produces such a translation that is based on two types of basic linguistic elements, that is, the morphological lexicon and the grammatical description. Most phases in the translation process can be implemented without major difficulties. The most difficult area in implementation is the choice among the readings that are grammatically identical but semantically different.

The system produces predictable translation, and each type of case is treated always in the same way. Therefore, the typing errors that usually constitute a large part in editing a manual translation are virtually absent in the rule-based translation system. Points where the system is not sure can be made to stand out for easy post-editing. Also un-translated words can be marked for later editing.

The translation system discussed here is, at least in principle, adaptable to any language. The adaptation is certainly easier to another Bantu language, because the structures of the source and target languages are similar. Two basic resources of target language are needed, that is, the lexicon that includes the proper names (about 2,750) and ordinary words (about 5,000) of the Bible, and the grammar of the language.

References

Bernth, A. and Gdanie, C., 2002. MTranslatability. *Machine Translation* 16:175–218.

Carl, M., Melero, M., Badia, T. Vandeghinste, V., Dirix, P., Schuurman, I., Markantonatou, S., Sofianopoulos, S., Vassiliou, M., Yannoutsou, O., 2008. METIS-II: low resource machine translation. *Machine Translation* 22:67–99.

Groves D. and Way, A., 2006. Hybrid data-driven models of machine translation. *Machine Translation* 19:301–323.

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. 1996. Disambiguation of morphological analysis in Bantu languages. *Proceedings of COLING 96*, pp. 568-573.

Hurskainen A. 2004a. Optimizing Disambiguation in Swahili. In *Proceedings of COLING 04, The 20th International Conference on Computational Linguistics, Geneva 23 27.8. 2004*. Pp. 254-260. http://www.njas.helsinki.fi

Hurskainen, A. 2004b. 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, Arvi, 2009. Two-Phase method in Morphological Analysis. *Technical Reports in Language Technology*. Report No 6. http://www.njas.helsinki.fi/salama

Karlsson, F. 1995. Designing a parser for unrestricted text. Karlsson, F. et al. (Eds.), *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*: 1-40. Berlin: Mouton de Gryuter.

Koskenniemi, K. 1983. *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*. Department of General Linguistics. University of Helsinki. Publication No. 11.

Mel'čuk, I. and Wanner, L., 2001. Towards a Lexicographic Approach to Lexical Transfer in Machine Translation (Illustrated by the German – Russian Language Pair). *Machine Translation*16:21–87.

Nyberg, E. H., Mitamura, T. and Carbonell, J. G., 1994. Evaluation metrics for knowledge-based machine translation. *Proceedings of the 15th conference on Computational linguistics - Volume 1*, pp. 95-99. Association for Computational Linguistics.

Pekar V., Mitkov, R., Blagoev, D. and Mulloni, A., 2006. Finding translations for low-frequency words in comparable corpora. *Machine Translation* 20:247–266.

Riezler, Stefan and Maxwell, John T. 2006. Grammatical machine translation. *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*, pp. 248–255. New York: Association for Computational 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.

Ten Hacken, P. 2001. Has There Been a Revolution in Machine Translation? *Machine Translation* 16: 1–19.

Appendix I

List of tags appearing in the examples.

[1-SG1-OBJ]	object prefix of verb, class 1, sg1 (Luganda)
[1-SG1-SP]	subject prefix of verb, class 1, sg1 (Luganda)
[1-SG2-OBJ]	object prefix of verb, class 1, sg2 (Luganda)
[1-SG2-SP]	subject prefix of verb, class 1, sg2 (Luganda)
[1-SG3-OBJ]	object prefix of verb, class 1, sg3 (Luganda)
[1-SG3-SP]	subject prefix of verb, class 1, sg3 (Luganda)
[10-PL-OBJ]	object prefix of verb, class 10, pl (Luganda)
[10-PL-SP]	subject prefix of verb, class 10, pl (Luganda)
[11-SG-OBJ]	object prefix of verb, class 11, sg (Luganda)
[11-SG-SP]	subject prefix of verb, class 11, sg (Luganda)
[12-SG-OBJ]	object prefix of verb, class 12, sg (Luganda)
[12-SG-SP]	subject prefix of verb, class 12, sg (Luganda)
[13-PL-OBJ]	object prefix of verb, class 13, pl (Luganda)
[13-PL-SP]	subject prefix of verb, class 13, pl (Luganda)
[14-PL-OBJ]	object prefix of verb, class 14, pl (Luganda)
[14-PL-SP]	subject prefix of verb, class 14, pl (Luganda)
[15-SG-OBJ]	object prefix of verb, class 15, sg (Luganda)
[15-SG-SP]	subject prefix of verb, class 15, sg (Luganda)
[16-SG-OBJ]	object prefix of verb, class 16, sg (Luganda)
[16-SG-SP]	subject prefix of verb, class 16, sg (Luganda)
[17-SG-OBJ]	object prefix of verb, class 17, sg (Luganda)
[17-SG-SP]	subject prefix of verb, class 17, sg (Luganda)
[18-SG-OBJ]	object prefix of verb, class 18, sg (Luganda)
[18-SG-SP]	subject prefix of verb, class 18, sg (Luganda)
[19-SG-OBJ]	object prefix of verb, class 19, sg (Luganda)
[19-SG-SP]	subject prefix of verb, class 19, sg (Luganda)
[2-PL1-OBJ]	object prefix of verb, class 2, pl1 (Luganda)
[2-PL1-SP]	subject prefix of verb, class 2, pl1 (Luganda)
[2-PL2-OBJ]	object prefix of verb, class 2, pl2 (Luganda)
[2-PL2-SP]	subject prefix of verb, class 2, pl2 (Luganda)
[2-PL3-OBJ]	object prefix of verb, class 2, pl3 (Luganda)
[2-PL3-SP]	subject prefix of verb, class 2, pl3 (Luganda)
[20-SG-OBJ]	object prefix of verb, class 20, sg (Luganda)
[20-SG-SP]	subject prefix of verb, class 20, sg (Luganda)
[22-PL-OBJ]	object prefix of verb, class 22, pl (Luganda)
[22-PL-SP]	subject prefix of verb, class 22, pl (Luganda)
[23-SG-OBJ]	object prefix of verb, class 23, sg (Luganda)
[23-SG-SP]	subject prefix of verb, class 23, sg (Luganda)
[3-SG-OBJ]	object prefix of verb, class 3, sg (Luganda)
[3-SG-SP]	subject prefix of verb, class 3, sg (Luganda)
[4-PL-OBJ]	object prefix of verb, class 4, pl (Luganda)
[4-PL-SP]	subject prefix of verb, class 4, pl (Luganda)
[5-SG-OBJ]	object prefix of verb, class 5, sg (Luganda)
[5-SG-SP]	subject prefix of verb, class 5, sg (Luganda)
[6-PL-OBJ]	object prefix of verb, class 6, pl (Luganda)
[6-PL-SP]	subject prefix of verb, class 6, pl (Luganda)

[7-SG-OBJ]	object prefix of verb, class 7, sg (Luganda)
[7-SG-SP]	subject prefix of verb, class 7, sg (Luganda)
[8-PL-OBJ]	object prefix of verb, class 8, pl (Luganda)
[8-PL-SP]	subject prefix of verb, class 8, pl (Luganda)
[9-SG-OBJ]	object prefix of verb, class 9, sg (Luganda)
[9-SG-SP]	subject prefix of verb, class 9, sg (Luganda)
a-2	genitive prefix, class 2 (Luganda)
wa-14	genitive prefix, class 14 (Luganda)
ya-8	genitive prefix, class 8 (Luganda)
a-23	genitive prefix, class 23 (Luganda)
a-6	genitive prefix, class 6 (Luganda)
wa-3	genitive prefix, class 3 (Luganda)
ya-4	genitive prefix, class 4 (Luganda)
a-12	genitive prefix, class 12 (Luganda)
wa-15	genitive prefix, class 15 (Luganda)
ya-7	genitive prefix, class 7 (Luganda)
wa-11	genitive prefix, class 11 (Luganda)
lya-5	genitive prefix, class 5 (Luganda)
a-1	genitive prefix, class 1 (Luganda)
a-9	genitive prefix, class 9 (Luganda)
a-10	genitive prefix, class 10 (Luganda)
[MAT28_19/]	unique code of the Bible verse
**CLB	clause boundary
1/2-PL	noun group 1/2, pl (Swahili)
2-PL2-OBJ	object prefix of verb, class 2, pl2 (Swahili)
2-PL2-SP	subject prefix of verb, class 2, pl2 (Swahili)
2-PL3-OBJ	object prefix of verb, class 2, pl3 (Swahili)
1/2-SG	noun group 1/2, sg (Swahili)
1-SG1-SP	subject prefix of verb, class 1, sg1 (Swahili)
1-SG3-SP	subject prefix of verb, class 1, sg3 (Swahili)
11-SG	class 11, sg (Swahili)
15-SG	class 15, sg (Swahili)
16-SG	class 16, sg (Swahili)
17-SG	class 17, sg (Swahili)
18-SG	class 18, sg (Swahili)
18-SG-SP	subject prefix of verb, class 18, sg (Swahili)
3/4-PL	noun group 3/4, pl (Swahili)
4-PL-REL	relative prefix of verb, class 4, pl (Swahili)
3/4-SG	noun group 3/4, sg (Swahili)
5/6-PL	noun group 5/6, pl (Swahili)
6-PL-OBJ	object prefix of verb, class 6, pl (Swahili)
6-PL-REL	relative prefix of verb, class 6, pl (Swahili)
5/6-SG	noun group 5/6, sg (Swahili)
6-PLSG-OBJ	object prefix of verb, class 2, plsg (Swahili)
6-PLSG-REL	relative prefix of verb, class 6, plsg (Swahili)
9/10-PL	noun group 9/10, pl (Swahili)
9/10-SG	noun group 9/10, sg (Swahili)
9-SG-REL	relative prefix of verb, class 9, sg (Swahili)
9/6-SG	noun group 9/6, sg (Swahili)
<aeon>	un-translated word
@<GN	head of genitive on the left

@ADVL	adverbial
@CC	coordinating connector
@FMAINVintr	finite intransitive main verb
@FMAINVtr+OBJ>	finite transitive main verb, object on the right
@FMAINVtr-OBJ>	finite transitive main verb, no overt object
@GCON	genitive connector
@OBJ	object of the clause
@SUBJ	subject of the clause
ADV	adverb
AG-PART	agentive particle
AN	animate
APPL	applicative verb extension
AR	Arabic origin
AUX-WA	auxiliary verb -wa
CAP	capital word
CARD	cardinal number
CAUS	causative verb extension
CC	coordinating conjunctive
COMMA	comma
COND-IF	conditional verb form, to be translated with 'if'
COND:ki	conditional verb form, with the marker -ki-
CONJ	conjunction
CONSEQ:ka-e	consequtive verb form, with the marker -ka- and verb-final e
ENG	English origin
EXCLAM	exclamation
GEN-CON	genitive connector
HUM	human
IMP	imperative verb form
IMP-PL2	imperative verb form, pl2
INF	infinitive verb form
INFMARK	infinitive marker
LINE-CODE	code of the verse
LOC	locative
LOC-16	locative, class 16
MALE	male
MOD-CAN	modal verb, to be translated with 'can'
MONOSLB	monosyllabic verb
N	noun
NARR:ka	narrative verb form, marker -ka-
NO-OBJ-GLOSS	no object prefix gloss allowed
NO-PRON-GLOSS	no pronoun gloss allowed
NO-SP-GLOSS	no subject prefix gloss allowed
NO-TO	no 'to' for an infinitive verb allowed
NOART	no article allowed
NUM	number
NUM-INFL	inflecting number
OBJ	object
PASS	passive verb form
PAST	past verb form
PERF:me	perfect verb form, marker -me-
PERS	Persian origin

PLACE	place
POSS	possessive pronoun
PREP	preposition
PRON	pronoun
PROPNAME	proper name
SBJN	subjunctive verb form
SV	intransitive verb
SVO	transitive verb
SVOO	ditransitive verb
TIME	time
TITLE	title
V	verb
V-BE	verb, with the gloss 'be'
VFIN	finite verb
[amba]	base stem of the verb