Technical Reports on Language Technology
Report No 18, 20014
http://www.njas.helsinki.fi/salama

Clashing and discontinuous multiword expressions[1]

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

Abstract

In Correct description of multiword expressions (MWE) is a challenging enterprise in machine translation (MT). Particularly problematic are cases where MWE candidates overlap and the correct candidates in that context must be selected. Another type of tricky cases is where one or more MWEs would be part of a larger MWE, if there were no context restrictions. These cases of MWEs will be discussed here.

Keywords: *machine translation, multiword expressions, disambiguation.*

1 Overlapping multiword expressions

Multiword expressions abound in normal text. It often happens that a word may be part in two overlapping MWE candidates. It is essential to control the environment for deciding, which MWE candidates are real MWEs in the context. Consider the example in (1).

(1)
Mtiririko wa matukio mbali mbali ya nyuma yanatudhihirishia kuwa kumekuwepo na mpango wa muda mrefu wa haya yanayotokea sasa hivi.

MWE candidates in the above sentence are: *mtirikiko wa matukio*, *mbali mbali*, *mbali ya*, *ya nyuma*, *kumekuwepo na*, *wa muda*, and *sasa hivi*. Out of these only *kumekuwepo na* and *sasa hivi* are unambiguously MWEs. In order to abbreviate the analysis results, below we will concentrate on the first part of the sentence, that is, *Mtiririko wa matukio mbali mbali ya nyuma*. The analysis result, without disambiguation, is in (2).

```
(2)
"<*mtiririko>"
      "mtiririko" N 3/4-SG { the } { flow , trickle } CAP
"<wa>"
      "wa" GEN-CON 3-SG { of }
      "wa" GEN-CON 11-SG { of }
```

[1] The report is issued under licence CC BY-NC

```
      "wa" GEN-CON 1-SG { of }
      "wa" GEN-CON 2-PL { of }
"<matukio>"
      "tukio" N 5/6-PL { the } { event , happening , occurrence }
"<mbali>"
      "bali" V SBJN VFIN 1-SG3-OBJ OBJ { him , her } z [bali] { be
unconcerned } SV
      "bali" V SBJN 2-PL2-SP VFIN { you } z [bali] { be
unconcerned } SV
      "mbali" ADV { far , afar , not right } PLACE
"<mbali>"
      "bali" V SBJN VFIN 1-SG3-OBJ OBJ { him , her } z [bali] { be
unconcerned } SV
      "bali" V SBJN 2-PL2-SP VFIN { you } z [bali] { be
unconcerned } SV
      "mbali" ADV { far , afar , not right } PLACE
"<ya>"
      "ya" GEN-CON 4-PL { of }
      "ya" GEN-CON 9-SG { of }
      "ya" GEN-CON 6-PL { of }
      "ya" GEN-CON 6-PLSG { of }
"<nyuma>"
      "nyuma" N 9/10-SG { the } { behind , past , former }
      "nyuma" N 9/10-PL { the } { behind , past , former }
      "uma" N 11/10-PL { the } { fork }
      "nyuma" ADV { behind }
```

We may still simplify the result by disambiguating without taking the MWEs into consideration (3).

```
(3)
"<*mtiririko>"
      "mtiririko" N 3/4-SG { the } { flow , trickle } CAP
"<wa>"
      "wa" GEN-CON 3-SG { of }
"<matukio>"
      "tukio" N 5/6-PL { the } { event , happening , occurrence }
"<mbali>"
      "mbali" ADV { far , afar , not right } PLACE
"<mbali>"
      "mbali" ADV { far , afar , not right } PLACE
"<ya>"
      "ya" GEN-CON 6-PL { of }
"<nyuma>"
      "nyuma" N 9/10-SG { the } { behind , past , former }
```

If we analyze the MWE candidates in isolation, without taking the context into consideration, we get the following result (4). Note that the words *mbali* and *ya* are part in two clashing MWEs.

```
(4)
```

```
"<*mtiririko_wa_matukio>"
      "mtiririko_wa_tukio" N 3/4-SG { the } { plot } INITCAP @SUBJ
"<mbali_mbali>"
      "mbali_mbali" ADJ { various }
"<mbali_ya>"
      "mbali_ya" PREP { apart from } @ADVL
"<ya_nyuma>"
      "ya_nyuma" ADJ { past }
```

Should *mtiririko wa matukio* be interpreted as a MWE in this context? It is not clear what the writer means, but the structure of the sentence shows that, if *mbali mbali* is interpreted as a MWE, whereby it is an adjective requiring the head noun in plural, its referent must be *matukio* and not *mtiririko*. Therefore *mtiririko wa matukio* cannot be a MWE.

The next MWE candidate *mbali mbali* does not have any constraint against its interpretatioon as a MWE.

The words *mbali ya* cannot be a MWE here, because the word *mbali* is immediately on the left and *nyuma* on the right.

The last MWE candidate *ya nyuma* is a true MWE. It is an adjective modifying the noun *matukio*.

Taking the above discussion into consideration we get the following analysis with two MWEs (5).

```
(5)
"<*mtiririko>"
      "mtiririko" N 3/4-SG { the } { flow } INITCAP @SUBJ
"<wa>"
      "wa" GEN-CON 3-SG { of } @GCON
"<matukio>"
      "tukio" N 5/6-PL { the } { event } @<GN
"<mbali_mbali>"
      "mbali_mbali" ADJ { various } @<NADJ
"<ya_nyuma>"
      "ya_nyuma" ADJ { past } @<NADJ
```

The final translation is in (6).

(6)
The flow of various past events

2 One or more MWEs being part of a larger MWE

It sometimes happens that a cluster that is itsel MWE is part of a larger MWE, without partially clashing with it. Consider example in (7).

```
(7)
"<upigaji>"
      "upigaji" N 11-SG { hitting }  >>>MW N >MW { voting }
      "upigaji" N 11-SG { beating }  >>>MW N >MW { voting }
```

```
"<kura>"
      "kura" N 9/10-SG { the } >>MW { :referendum }
      "kura" N 9/10-PL { the } >>MW { :referendum }
"<ya>"
      "ya" GEN-CON 4-PL { of }
      "ya" GEN-CON 9-SG { of }
      "ya" GEN-CON 6-PL { of }
      "ya" GEN-CON 6-PLSG { of }
"<maoni>"
      "oni" N 5/6-PL { the } { view }
      "oni" N 5/6-PL { the } { opinion }
```

If *upigaji kura ya maoni* is considered as a MWE, there are two MWEs inside this cluster, *upigaji kura* 'voting' and *kura ya maoni* 'referendum'. In this case, the whole word cluster is marked as a MWE with the tag >>>MW, but another rule with the tag >MW has replaced the gloss given by the first rule. Therefore the gloss 'voting'.

The problem is solved by correct rule ordering using the 'longest first' principle. The application of shorter rules is then restricted so that if the reading or its preceding reading(s) have a MWE tag, the rule should not be applied. When we modify the rules accordingly, we get the result as in (8).

```
(8
"<upigaji>"
      "upigaji" N 11-SG >>>MW { :referendum :voting }
"<kura>"
      "kura" N 9/10-SG { the } { vote }
      "kura" N 9/10-SG { the } { ballot }
      "kura" N 9/10-PL { the } { vote }
      "kura" N 9/10-PL { the } { ballot }
"<ya>"
      "ya" GEN-CON 4-PL { of }
      "ya" GEN-CON 9-SG { of }
      "ya" GEN-CON 6-PL { of }
      "ya" GEN-CON 6-PLSG { of }
"<maoni>"
      "oni" N 5/6-PL { the } { view }
      "oni" N 5/6-PL { the } { opinion }
```

In other contexts the shorter rules apply. Consider examples (9 and 10).

```
(9
"<wakati>"
      "wakati" N 11/10-SG { the } { time }  TIME
      "wakati" N 11/10-SG { the } { period of :time }  TIME
      "wakati" N 11/10-SG { the } { point of :time }  TIME
      "wakati" ADV { while }
      "wakati" ADV { when }
"<wa>"
      "wa" GEN-CON 3-SG { of }
      "wa" GEN-CON 11-SG { of }
```

```
      "wa" GEN-CON 1-SG { of }
      "wa" GEN-CON 2-PL { of }
"<kura>"
      "kura" N 9/10-SG { the } >>MW { :referendum }
      "kura" N 9/10-PL { the } >>MW { :referendum }
"<ya>"
      "ya" GEN-CON 4-PL { of }
      "ya" GEN-CON 9-SG { of }
      "ya" GEN-CON 6-PL { of }
      "ya" GEN-CON 6-PLSG { of }
"<maoni>"
      "oni" N 5/6-PL { the } { view }
      "oni" N 5/6-PL { the } { opinion }

(10)
"<*wakati>"
      "wakati" N 11/10-SG { the } { time }  TIME CAP
      "wakati" N 11/10-SG { the } { period of :time }  TIME CAP
      "wakati" N 11/10-SG { the } { point of :time }  TIME CAP
      "wakati" ADV { while }  CAP
      "wakati" ADV { when }  CAP
"<upigaji>"
      "upigaji" N 11-SG N >MW { voting }
"<kura>"
      "kura" N 9/10-SG { the } { vote }
      "kura" N 9/10-SG { the } { ballot }
      "kura" N 9/10-PL { the } { vote }
      "kura" N 9/10-PL { the } { ballot }
"<ulipoanza>"
      "anza" V 1-SG2-SP VFIN { you } PAST 16-SG-REL { when } z
[anza] { begin } SVO
      "anza" V 1-SG2-SP VFIN { you } PAST 16-SG-REL { when } z
[anza] { establish } SVO
      "anza" V 1-SG2-SP VFIN { you } PAST 16-SG-REL { where } z
[anza] { begin } SVO
      "anza" V 1-SG2-SP VFIN { you } PAST 16-SG-REL { where } z
[anza] { establish } SVO
      "anza" V 3-SG-SP VFIN { it } PAST 16-SG-REL { when } z
[anza] { begin } SVO
      "anza" V 3-SG-SP VFIN { it } PAST 16-SG-REL { when } z
[anza] { establish } SVO
      "anza" V 3-SG-SP VFIN { it } PAST 16-SG-REL { where } z
[anza] { begin } SVO
      "anza" V 3-SG-SP VFIN { it } PAST 16-SG-REL { where } z
[anza] { establish } SVO
      "anza" V 11-SG-SP VFIN { it } PAST 16-SG-REL { when } z
[anza] { begin } SVO
      "anza" V 11-SG-SP VFIN { it } PAST 16-SG-REL { when } z
[anza] { establish } SVO
      "anza" V 11-SG-SP VFIN { it } PAST 16-SG-REL { where } z
[anza] { begin } SVO
```

```
      "anza" V 11-SG-SP VFIN { it } PAST 16-SG-REL { where } z
[anza] { establish } SVO
```

3 MWEs as verbs

Swahili has a large number of such verbs, the meaning of which is expressed with the verb and its object. Examples include *piga hatua* 'advance', 'take steps', *piga picha* 'photograph', 'take a photo', and *chapa kazi* 'work hard'. The normal use of these idiomatic expressions is that the verb is followed by the object (11).

```
(11)
"<*serikali>"
      "serikali" N 9/10-SG { the } { government } INITCAP @SUBJ
"<ilipiga_hatua>"
      "piga_hatua" V 9-SG-SP VFIN NO-SP-GLOSS PAST z [piga] SVO
ACT V { advance } @FMAINVtr-OBJ>
"<na>"
      "na" CC { and } @CC
"<mpiga_picha>"
      "mpiga_picha" N 1/2-SG HUM { the } HUM N { photographer }
@SUBJ
"<alipiga_picha>"
      "piga_picha" V 1-SG3-SP VFIN NO-SP-GLOSS PAST z [piga] SVO
ACT V { photograph } @FMAINVtr-OBJ>
```

The final traslation is in (12)

(12)
The government advanced and the photographer photographed

These MWEs can also be used in relative constructions, so that the object precedes the verb. Such constructions cannot be translated with a simple English verb (13).

```
(13)
"<*hatua>"
      "hatua" N 9/10-SG { the } MW-CAND { step }  CAP
      "hatua" N 9/10-SG { the } MW-CAND { pace in :walking }  CAP
      "hatua" N 9/10-SG { the } MW-CAND { progress }  CAP
      "hatua" N 9/10-PL { the } MW-CAND { step }  CAP
      "hatua" N 9/10-PL { the } MW-CAND { pace in :walking }  CAP
      "hatua" N 9/10-PL { the } MW-CAND { progress }  CAP
"<zilizopigwa>"
      "pigwa" V 10-PL-SP VFIN { they } PAST 10-PL-REL { which } z
[piga] { hit } SVO ACT PASS SUB-REL
      "pigwa" V 10-PL-SP VFIN { they } PAST 10-PL-REL { which } z
[piga] { beat } SVO ACT PASS SUB-REL
      "pigwa" V 10-PL-SP VFIN { they } PAST 10-PL-REL { which } z
[piga] { take } SVO ACT PASS SUB-REL
"<na>"
```

```
      "na" CC { and }
      "na" AG-PART { by }
      "na" PREP { with }
      "na" NA-POSS { of }
      "na" ADV NOART { past }
      "na" ADV { also }
"<serikali>"
      "serikali" N 9/10-SG { the } { government }
      "serikali" N 9/10-PL { the } { government }
```

The verb *[piga]* above has three interpretations, 'hit', 'beat', and 'take'. The two first ones are default meanings, and the third one 'take' is used only in some idioms. With disambiguation rules the correct meaning can be selected (14).

```
(14)
"<*hatua>"
      "hatua" N 9/10-PL { the } MW-CAND { step } INITCAP @SUBJ+rel
"<zilizopigwa>"
      "pigwa" V 10-PL-SP VFIN { they } PAST 10-PL-REL { which } z
[piga] { take } SVO ACT PASS SUB-REL @FMAINVtr-OBJ>
"<na>"
      "na" AG-PART { by } @ADVL
"<serikali>"
      "serikali" N 9/10-PL { the } { government } @SUBJ
"<na>"
      "na" CC { and } @CC
"<picha>"
      "picha" N 9/10-PL { the } { picture } @SUBJ
"<zilizopigwa>"
      "pigwa" V 10-PL-SP VFIN NO-SP-GLOSS PAST 10-PL-REL { which }
z [piga] { take } SVO ACT PASS SUB-REL @FMAINVtr-OBJ>
"<na>"
      "na" AG-PART { by } @ADVL
"<mpiga_picha>"
      "mpiga_picha" N 1/2-SG HUM { the } HUM N { photographer }
@AG
```

The final translation is in (15).

(15)
Steps which were taken by the governments and the pictures which were taken by the photographer

4 Discontinuous MWEs

MWEs may be discontinuous in allowing other words between the members of the MWE. Consider the example in (16).

```
(16)
```

```
"<*moto>"
      "moto" N 3/4-SG { the } { fire } CAP
"<huo>"
      "huo" PRON DEM :hV ASS-OBJ 3-SG { this }
      "huo" PRON DEM :hV ASS-OBJ 3-SG { that }
      "huo" PRON DEM :hV ASS-OBJ 11-SG { this }
      "huo" PRON DEM :hV ASS-OBJ 11-SG { that }
"<ulianza>"
      "anza" V 1-SG2-SP VFIN { you } PAST z [anza] { begin } SVO
      "anza" V 1-SG2-SP VFIN { you } PAST z [anza] { establish }
SVO
      "anza" V 3-SG-SP VFIN { it } PAST z [anza] { begin } SVO
      "anza" V 3-SG-SP VFIN { it } PAST z [anza] { establish } SVO
      "anza" V 11-SG-SP VFIN { it } PAST z [anza] { begin } SVO
      "anza" V 11-SG-SP VFIN { it } PAST z [anza] { establish }
SVO
"<katika>"
      "katika" V IMP VFIN z [kata] { be :cut } SV STAT PREFR
      "katika" V IMP VFIN z [kata] { cut } SV STAT
      "katika" V IMP VFIN z [kata] { hew } SV STAT
      "katika" V IMP VFIN z [kata] { mince } SV STAT
      "katika" V <kwisha z [kata] { be :cut } SV STAT PREFR
      "katika" V <kwisha z [kata] { cut } SV STAT
      "katika" V <kwisha z [kata] { hew } SV STAT
      "katika" V <kwisha z [kata] { mince } SV STAT
      "tika" V NARR-COLLOQ:ka-a VFIN z [tia] { put } SV STAT
      "tika" V NARR-COLLOQ:ka-a VFIN z [tia] { pour } SV STAT
      "tika" V NARR-COLLOQ:ka-a VFIN z [tia] { bring about } SV
STAT
      "katika" PREP { in }
      "katika" PREP { at }
      "katika" PREP { on }
      "katika" PREP { from }
      "katika" IMP VFIN z [kata] { be :cut } SV STAT PREFR
      "katika" IMP VFIN z [kata] { cut } SV STAT
      "katika" IMP VFIN z [kata] { hew } SV STAT
      "katika" IMP VFIN z [kata] { mince } SV STAT
      "katika" <kwisha z [kata] { be :cut } SV STAT PREFR
      "katika" <kwisha z [kata] { cut } SV STAT
      "katika" <kwisha z [kata] { hew } SV STAT
      "katika" <kwisha z [kata] { mince } SV STAT
      "tika" NARR-COLLOQ:ka-a VFIN z [tia] { put } SV STAT
      "tika" NARR-COLLOQ:ka-a VFIN z [tia] { pour } SV STAT
      "tika" NARR-COLLOQ:ka-a VFIN z [tia] { bring about } SV STAT
"<nyumba>"
      "nyumba" N 9/10-SG { the } PLACE x>>>MW { :guest house }
      "nyumba" N 9/10-PL { the } PLACE x>>>MW { :guest house }
"<hiyo>"
      "hiyo" PRON DEM :hV ASS-OBJ 4-PL { these }
      "hiyo" PRON DEM :hV ASS-OBJ 6-PL { these }
      "hiyo" PRON DEM :hV ASS-OBJ 6-PLSG { this }
```

```
      "hiyo" PRON DEM :hV ASS-OBJ 9-SG { this }
"<ya>"
      "ya" GEN-CON 4-PL { of }
      "ya" GEN-CON 9-SG { of }
      "ya" GEN-CON 6-PL { of }
      "ya" GEN-CON 6-PLSG { of }
"<kulala>"
      "lala" V INF { to } z [lala] { sleep } PERF-STATE SV
      "lala" V INF { to } z [lala] { lie down } PERF-STATE SV
      "lala" V INF { to } z [lala] { have sexual intercourse }
PERF-STATE SV
      "lala" N 15-SG z [lala] { sleep } PERF-STATE SV
      "lala" N 15-SG z [lala] { lie down } PERF-STATE SV
      "lala" N 15-SG z [lala] { have sexual intercourse } PERF-
STATE SV
      "lala" V INF MOD-CAN z [lala] { sleep } PERF-STATE SV
      "lala" V INF MOD-CAN z [lala] { lie down } PERF-STATE SV
      "lala" V INF MOD-CAN z [lala] { have sexual intercourse }
PERF-STATE SV
      "lala" V INF NO-TO z [lala] { sleep } PERF-STATE SV
      "lala" V INF NO-TO z [lala] { lie down } PERF-STATE SV
      "lala" V INF NO-TO z [lala] { have sexual intercourse }
PERF-STATE SV
"<wageni>"
      "mgeni" N 1/2-PL HUM { the } { guest }
      "mgeni" N 1/2-PL HUM { the } { visitor }
      "mgeni" N 1/2-PL HUM { the } { stranger }
      "mgeni" N 1/2-PL HUM { the } { foreigner }
      "mgeni" N 1/2-PL HUM { the } { alien }
      "wage" N 11-SG { in } { itchy :plant hairs } LOC
      "wage" N 11-SG { on } { itchy :plant hairs } LOC
      "geni" ADJ A-INFL 2-PL { foreign }
      "geni" ADJ A-INFL 2-PL { strange }
"<.$>"
      ".$" { .$ } **CLB
```

Th MWE is *nymba ya kulala wageni*, but it allows a word, normally a demonstrative or possessive pronoun, or an adjective or numeral, after the word *nyumba*. The tag `x>>>MW` on *nyumba* indicates that it is part of the MWE, the following word is not part of it, but the second, third and fourth word is. The result after marking other members of the MWE is in (17).

```
(17)
"<<s>>"
      "<s>" { <s> }
"<*moto>"
      "moto" N 3/4-SG { the } { fire } CAP
"<huo>"
      "huo" PRON DEM :hV ASS-OBJ 3-SG { this }
"<ulianza>"
      "anza" V 3-SG-SP VFIN { it } PAST z [anza] { begin } SVO
```

```
"<katika>"
      "katika" PREP { in }
"<nyumba>"
      "nyumba" N 9/10-SG { the } PLACE x>>>MW { :guest house }
"<hiyo>"
      "hiyo" PRON DEM :hV ASS-OBJ 9-SG { this }
"<ya>"
      "ya" MW<x>>
"<kulala>"
      "lala" MW<x<>
"<wageni>"
      "mgeni" MW<x<<
"<.$>"
      ".$" { .$ } **CLB
```

Translation.

(14)
This fire began in this guest house.

Another example of a discontinuous MWE is in (18). It allows two numerals after the first member of the MWE.

```
(18)
"<*zilinunuliwa>"
      "nunuliwa" V 10-PL-SP VFIN { they } PAST z [nunua] { buy }
SVO PASS CAP
"<pia>"
      "pia" ADV { also }  PREFR
"<na>"
      "na" AG-PART { by }
"<*pap>"
      "*pap" <Heur> PROPNAME { *pap } CAP
"<kwa>"
      "kwa" PREP { for }
"<dola>"
      "dola" N 9/10-PL { the } xx>>MW { *:american dollars }
"<milioni>"
      "milioni" NUM NUM-UNINFL CARD { million }
"<75>"
      "75" NUM { 75 }
"<za>"
      "za" MW<xx>
"<*marekani>"
      "*marekani" MW<xx<
"<.$>"
      ".$" { .$ } **CLB
"<zilinunuliwa>"
      "nunuliwa" V 10-PL-SP VFIN { they } PAST z [nunua] { buy }
SVO PASS
```

```
"<pia>"
      "pia" ADV { also }  PREFR
"<na>"
      "na" AG-PART { by }
"<*pap>"
      "*pap" <Heur> PROPNAME { *pap } CAP
"<kwa>"
      "kwa" PREP { for }
"<dola>"
      "dola" N 9/10-PL { the } xx>>MW { *:american dollars }
"<milioni>"
      "milioni" NUM NUM-UNINFL CARD { million }
"<75>"
      "75" NUM { 75 }
"<za>"
      "za" MW<xx>
"<*marekani>"
      "*marekani" MW<xx<
"<.$>"
      ".$" { .$ } **CLB
```

Translation.

(19)
They were bought also by Pap for 75 million American dollars.

5 Several words inside the MWE cluster

Sometimes there are more than one or two extra words within the MWE cluster. In that case we have to scan to the right to find the other members of the MWE and mark the first member (17). Note that the first member nyumba is marked with `*>>>MW` indicating that the other members are on the right, but their distance is not defined.

```
(20)
"<*nilikaa>"
      "kaa" V 1-SG1-SP VFIN { *i } PAST z [kaa] { stay } SVO CAP
"<katika>"
      "katika" PREP { in }
"<nyumba>"
      "nyumba" N 9/10-SG { the } PLACE *>>>MW { :guest house }
"<yetu>"
      "etu" PRON POSS 9-SG PL1 { our }
"<nzuri>"
      "zuri" ADJ A-INFL 9-SG { good }
"<ile>"
      "ile" PRON DEM :le 9-SG { that }
"<ya>"
      "ya" GEN-CON 9-SG { of }
"<kulala>"
      "lala" N 15-SG z [lala] { sleep } PERF-STATE SV
```

```
"<wageni>"
      "mgeni" N 1/2-PL HUM { the } { guest }
"<.$>"
      ".$" { .$ } **CLB
```

Then we mark the other members of the structure using a heuristic rule.

```
(21)
"<*nilikaa>"
      "kaa" V 1-SG1-SP VFIN { *i } PAST z [kaa] { stay } SVO
INITCAP @FMAINVtr-OBJ>
"<katika>"
      "katika" PREP { in } @ADVL
"<nyumba>"
      "nyumba" N 9/10-SG { the } PLACE *>>>MW { :guest house } @<P
"<yetu>"
      "etu" PRON POSS 9-SG PL1 { our } @GCON
"<nzuri>"
      "zuri" ADJ A-INFL 9-SG { good } @<NADJ
"<ile>"
      "ile" PRON DEM :le 9-SG { that } @OBJ
"<ya>"
      "ya" MW<*>>
"<kulala>"
      "lala" MW<*>>
"<wageni>"
      "mgeni" MW<*>>
"<.$>"
      ".$" { .$ } **CLB
```

Translation.

(22)
I stayed in that our good guest house.

We see that although the MWE is fragmented in the source language, it is not allowed to be fragmented in the TL. We cannot put other words within the MWE 'guest house'. These kinds of fragmented MWEs are common in Swahili.

We will make one more test with a longer word string in between. It has five extra words within the MWE cluster (23).

```
(23)
"<*nilikaa>"
      "kaa" V 1-SG1-SP VFIN { *i } PAST z [kaa] { stay } SVO CAP
"<katika>"
      "katika" PREP { in }
"<nyumba>"
      "nyumba" N 9/10-SG { the } PLACE *>>>MW { :guest house }
"<yetu>"
```

```
      "etu" PRON POSS 9-SG PL1 { our }
"<nzuri>"
      "zuri" ADJ A-INFL 9-SG { good }
"<na>"
      "na" CC { and }
"<ghali>"
      "ghali" ADJ A-UNINFL { expensive }
"<ile>"
      "ile" PRON DEM :le 9-SG { that }
"<ya>"
      "ya" GEN-CON 9-SG { of }
"<kulala>"
      "lala" N 15-SG z [lala] { sleep } PERF-STATE SV
"<wageni>"
      "mgeni" N 1/2-PL HUM { the } { guest }
"<.$>"
      ".$" { .$ } **CLB
```

The other members of the MWE are marked (24).

```
(24)
"<*nilikaa>"
      "kaa" V 1-SG1-SP VFIN { *i } PAST z [kaa] { stay } SVO
INITCAP @FMAINVtr-OBJ>
"<katika>"
      "katika" PREP { in } @ADVL
"<nyumba>"
      "nyumba" N 9/10-SG { the } PLACE *>>>MW { :guest house } @<P
"<yetu>"
      "etu" PRON POSS 9-SG PL1 { our } @GCON
"<nzuri>"
      "zuri" ADJ A-INFL 9-SG { good } @<NADJ
"<na>"
      "na" CC { and } @CC
"<ghali>"
      "ghali" ADJ A-UNINFL { expensive }
"<ile>"
      "ile" PRON DEM :le 9-SG { that } @OBJ
"<ya>"
      "ya" MW<*>>
"<kulala>"
      "lala" MW<*>>
"<wageni>"
      "mgeni" MW<*>>
"<.$>"
      ".$" { .$ } **CLB
```

Translation.

(25)
I stayed in that our good and expensive guest house.

Wen can use also plural forms of MWEs if they are nouns. The same rules apply also to them. Consider the example in (26).

```
(26)
"<*nilikaa>"
      "kaa" V 1-SG1-SP VFIN { *i } PAST z [kaa] { stay } SVO
INITCAP @FMAINVtr-OBJ>
"<katika>"
      "katika" PREP { in } @ADVL
"<nyumba>"
      "nyumba" N 9/10-PL { the } PLACE *>>>MW { :guest house } @<P
"<zetu>"
      "etu" PRON POSS 10-PL PL1 { our } @GCON
"<nzuri>"
      "zuri" ADJ A-INFL 10-PL { good } @<NADJ
"<na>"
      "na" CC { and } @CC
"<ghali>"
      "ghali" ADJ A-UNINFL { expensive }
"<zile>"
      "zile" PRON DEM :le 10-PL { those } @OBJ
"<za>"
      "za" MW<*>>
"<kulala>"
      "lala" MW<*>>
"<wageni>"
      "mgeni" MW<*>>
"<.$>"
      ".$" { .$ } **CLB
```

Translation.

(27)
I stayed in those our good and expensive guest houses.

6 Discussion

Swahili MWEs behave in various ways, and each type must be treated accordingly. In compound nouns, the head word must be selected as target and the existing grammatical information on it must be retained, so that it can be used for producing the correct surface form in TL. There is often a danger of clashing or overlapping MWEs. Rules must be designed carefully for avoiding wrong concatenation. Discontinuous MWEs are a problem of its own. Also they must be treated carefully to avoid wrong concatenation. Especially the MWEs that allow several other words in between are problematic.