

Accurate information retrieval from large corpora: Non-extended Swahili monosyllabic verbs¹

Arvi Hurskainen
Department of World Cultures, Box 59
FIN-00014 University of Helsinki, Finland
arvi.hurskainen@helsinki.fi
DOI: 10.13140/RG.2.2.14418.89281

Abstract

Accurate information retrieval in strongly inflecting languages is problematic, because the inflected forms of words make the formulation of the search key difficult. Often the result has weaknesses in recall or precision, or in both. There is strong motivation for such a search system, where we have access also to the base form of the word, and the search could be made on the basis of the base form instead of the surface form. This approach would require the analysis of the source text and further modifications for making the text suitable for search.

Small and medium-size texts could be handled in this way, but it is not feasible to use this method for large text corpora. The analysis and disambiguation routines would take too much time for making the method suitable for large texts.

There is, however, a solution to the problem. The search can be made in two phases, so that first we use the normal string search and get the sentences that contain the desired words, plus many such sentences that we do not want. In the second phase we process the sentences into such a format, where each word has also its base form included. Now the search is made using the base form as key. The result will have only those sentences that we want to have.

I test the method with Swahili mono-syllabic verbs, because they are very difficult to handle in search tasks. Swahili has a small set of monosyllabic verbs, which seem to belong to the very basic core of the language. Because the verbs may take prefixes and suffixes, it is not always easy to judge whether the word is a verb or not. Some monosyllabic verbs also have more than one meaning, which poses a challenge to disambiguation in translating Swahili to another language. The mono-syllabic verbs have also extended forms, some of which are not any more monosyllabic. They behave in a different way than the mono-syllabic versions, which also causes a problem in describing the verbs morphologically.

In this report I restrict to testing how the non-extended monosyllabic verbs can be retrieved from large text corpora. I exclude such forms, which have the final character *e* (subjunctive) or *i* (present tense negative). I also exclude extended forms.

Key Words: *information retrieval, language analysis.*

¹ The report is issued under licence CC BY-NC

1 Introduction

Among the core verbs in Swahili are mono-syllabic verbs. They belong to the very old layer of words in language. They also occur in text frequently, which is why they should be handled faultlessly in language technology. The fact that the stable element in verb is just a single consonant, or in some cases a cluster of consonants, makes the identification of mono-syllabic verbs difficult.

Because mono-syllabic verbs in text have very little stable material, that is, only the stem consonant or consonant cluster, direct string search is not ideal method because of excessive noise. The result may be either over- or under-productive, but only seldom accurate.

In this report I describe a method of retrieving monosyllabic verbs accurately, although the source text is not tagged. In tests I use the non-tagged corpus of 25 million Swahili words.

The search is made in two phases. First, the normal string search is made by using such search keys, which cover all forms of the verb. Such keys will find all true hits but also many such cases, which are not monosyllabic verbs.

If the context of the hit is sentence-long, the result will consist of true sentences, which are suitable for disambiguation.

After the first round we have a set of sentences, which contains all true hits but also many unwanted sentences.

In the second phase we analyse and disambiguate the sentences and then we retrieve the verbs using the base form of the verb as search key. It is also possible to use the root form of the verb as search key, because the analysis contains the base form and root form of each verb.

If the source text is very big, as the Swahili corpus is, it is likely that the text after the first search round is still too big to be analysed in fly. It may be advisable to remove such wrong hits, which are easily removable without risking the true hits. I will show the process below.

2 Search in two phases

Accurate information retrieval from large text corpora needs a two-phase process. First we retrieve all those sentences, which contain the targeted words. The result should have all the targeted words. In practice, however, the result includes also such sentences, which match with the search string, but the matched word does not belong to the category of intended matches. In other words, the result has all the sentences that we want, but also such sentences that we do not want.

The text that we will now deal with is much smaller than the original corpus, often just a small fraction of it. The size of the resulting text depends on what we are searching for. For example, if we search all occurrences of the verb *ku-wa* (to be), it is likely that the resulting text is still too large to be handled in the two-phase process. Such search tasks are, however, exceptional, and in most search tasks the method that we are discussing here is suitable.

Below I will test the proposed search method with some monosyllabic verbs.

2.1 The verb *ku-nya* (to fall, to drop, to shit)

With the search key 'nya ' we get 75802 hits from the Swahili corpus, far too many for using the advanced search method.² The reason is that in addition to the true hits we get many wrong hits, such as *ku-fanya* (to do).

We can solve the problem by removing from the first search result such common hits that are definitely wrong.

In (1) we see the first 30 hits with the search key 'nya '.

(1)

```
$ cat-allmat | kwic 'nya ' | head -30
naye kama David Onyango, amefanya tukio la kushangaza ambalo
a lilitokea kwenye Ikulu ya Kenya Desemba 22 mwaka huu kweny
wakiwa wanamwangalia bila kufanya lolote.
. Onyango ambaye baada ya kufanya tukio hilo alishikiliwa na
cho cha hatari ambacho kilimfanya akeshe usiku huo ndani ya
"Alifanya maamuzi yote makubwa peke
imeripotiwa kukamatwa kwa Wakenya wawili, kuhusiana na wizi
sa hizo za CRDB, ikidaiwa Wakenya hao walikutwa na kiasi kik
Wakenya hao ambao bado wanashikili
bao bado wanashikiliwa huko Kenya ni Charles Njoka, 36, na K
kipindupindu wakisema amewafanya washerehekee New Year na n
Licha ya kuzuiwa kufanya biashara za vyakula na vin
ta iliyopita, na kwamba kinafanya pia kazi ya kupiga vita ma
Akasema wamekuwa wakifanya kazi hiyo katika maeneo mb
"Kwa kweli tunafanya kazi hii katika mazingira
nye vituo mbalimbali vinavyofanya kazi za Ukimwi.
kiwa na mfadhili tutaweza kufanya vizuri kazi yetu hii ya ku
gamoto ambayo itawasaidia kufanya kazi zao kwa kujiamini zai
isema kuwa TOC inatarajia kufanya mipango ili kukutana na vi
uwa mgogoro huo huenda ukawafanya viongozi wa FAT kushindwa
ma kwamba wanataka kwanza kufanya uchunguzi wa kina juu ya k
iliana na Polisi ili waje kufanya uchunguzi wao, lakini hadi
kwa Polisi hao kufika leo kufanya uchunguzi.
Akasema watafanya hivyo kama vile walivyokub
fanyabiashara kutoka nchini Kenya ambao hununua machungwa ya
kwamba dalali yeyote atakayefanya kinyume na amri hiyo atach
oyo wa kuungana kama walivyofanya huko Kenya.
Kenya imemaliza uchaguzi wake hi
enya, ni kwamba wananchi wa Kenya waliwalazimisha wapinzani
is Mkapu akaombwa akubali kufanya mabadiliko ya Katiba, Tume
```

The true hits are so rare that this sample does not contain a single one. When we remove obvious wrong hits using the script in (2), we will get a much reduced text, with only 141 sentences

² The search key 'nya ' excludes such forms as *nyi* (present tense negative) and *nye* (subjunctive). However, these forms are not common.

(2)

```
$ cat-allmat | egrep 'nya ' | egrep -v  
"(fa|Fa|Ke|ke|o|sa|o|ya|e|pa|wa|ka|ga|Sa|a|ha|ch|hu|ia|mi|ng.u|bi|Nku)nya  
"
```

Part of the modified result is in (3). We see that now there are many true hits but also wrong hits. The result was reformulated using `kwic`, because `egrep` does not do aligning.

(3)

```
Mshitakiwa alibaki kinya huku amejinamia kama anay  
iti, huku Rajabu Kigundula akinya kuwa nafasi ya Ukatibu wa  
yumba ya kulala wageni ya Turunya iliyoko Mwenge Jijini, Wil  
INTERVIEWEE(S): 1. Mgunya Mwinyi (MG), 2. Maalim Haj  
wa sehemu husema hu... si Wagunya na kadhalika, wapo.  
ima, ee. Haya, ikawa vua inakunya kunya ngia ga uko kati.  
e. Haya, ikawa vua inakunya kunya ngia ga uko kati.  
a MV M Bali aaah kwa vua inakunya uzuri.
```

```
    Ka MV M Inakunya vyema.  
    MB M Wazee wa zamani, matunya hamna, watu wengi sasa haw  
dua kwamba, nafikiri mnao Wagunya kiasi.
```

```
    kayumila maneno pannembele kunya baaya linamba libali kulik  
pane yungula kwaa fisi nkwa kunya mavi nkutukuta mpaka kulyu  
X1 F Akenda kunya haja chumbani kule akaja k  
ya haja chumbani kule akaja kunya ukumbini akenda kunya bara  
akaja kunya ukumbini akenda kunya barazani.
```

```
    X1 F Akenda kunya haja. Chumbani kule. Akaja  
haja. Chumbani kule. Akaja kunya ukumbini, akenda kunya bar  
kaja kunya ukumbini, akenda kunya barazani, jamaa wale. Na y  
isi kuzaliwa kwetu kwa mamungunya (kicheko),  
oni kwa giza. Hapo ilianzia kunya mvua asubuhi, mvua iliilie  
siku nyingine. Anye, akesha kunya zile zile kata-- kanapakus  
ajili ya tamaa yake ya kuibiginya nchi hiyo hadi kufa.  
stan na India ambavyo vimejikunya katika mpaka wao wa Kashmi  
ombi ambalo limekutana na ukimnya mzito kutoka mataifa tajir  
ya nchi hiyo kulaumiwa vikali nya Jumuiya ya Ulaya kwa kukan  
ubwa wa mikate yanapokea vinyunya vya mikate iliyogandikwa k  
na Bemba alisema waasi wa chamnya chake cha MLC walikuwa wan
```

Now the text has such size that we can analyse and disambiguate it and process further, so that it is suitable for accurate search.

If we only need to find the hits without context, we can search from the disambiguated text, because it contains base forms of words (4).

(4)

```
$ cat-allmat | egrep 'nya ' | egrep -v  
"(fa|Fa|Ke|ke|o|sa|o|ya|e|pa|wa|ka|ga|Sa|a|ha|ch|hu|ia|mi|ng.u|bi|Nku)nya  
" | atta | swasent | wordlist | dis | perl -pe 's/\>'( )?\n/>' /gm' |  
egrep '"<' | perl -pe 's/\t//gm' | egrep '"nya"  
"<akinya>" "nya" V 1-SG3-SP VFIN NO-SP-GLOSS COND-IF PR:na { when } z  
[nya] { shit } SV @FMAINVtr-OBJ>  
"<inakunya>" "nya" V 9-SG-SP VFIN { it } PR:na 1-SG2-OBJ OBJ NO-OBJ-GLOSS  
z [nya] { shit } SV @FMAINVtr+OBJ>  
"<kunya>" "nya" N 15-SG z [nya] { shit } SV @OBJ
```

```
"<inakunya>" "nya" V 9-SG-SP VFIN { it } PR:na 1-SG2-OBJ OBJ NO-OBJ-GLOSS
z [nya] { shit } SV @FMAINVtr+OBJ>
"<*inakunya>" "nya"
"<kunya>" "nya" N 15-SG z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" N 15-SG z [nya] { shit } SV @PCOMPL-S
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" V INF { to } z [nya] { rain } SV @-FMAINV-n
"<*anye>" "nya" V SBJN 1-SG3-SP VFIN { she } z [nya] { *shit } SV CAP
@FMAINVtr-OBJ>
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<nya>" "nya" <Heur> N 9/10-SG { nya } @OBJ
"<alinya>" "nya" V 1-SG3-SP VFIN NO-SP-GLOSS PAST z [nya] { shit } SV
@FMAINVtr+OBJ>
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<kunya>" "nya" N 15-SG z [nya] { shit } SV @<P
"<tuliokunya>" "nya" V 2-PL1-SP VFIN NO-SP-GLOSS PAST 2-PL-SUB-REL { who
} INFMARK z [nya] { drop } SVO MONOSLB @FMAINVtr-OBJ>
"<nya>" "nya" <Heur> N 9/10-SG { nya } @OBJ
"<kunya>" "nya" V INF { to } z [nya] { shit } SV @-FMAINV-n
"<*ikinya>" "nya" V 4-PL-SP VFIN NO-SP-GLOSS COND-IF PAST { when } z
[nya] { *shit } SV CAP @FMAINVtr-OBJ>
"<*kunya>" "nya" V INF { to } z [nya] { *shit } SV CAP @-FMAINV-n
"<nya>" "nya" <Heur> N 9/10-SG { nya } @OBJ
"<nya>" "nya" <Heur> N 9/10-SG { nya } @SUBJ
"<anakunya>" "nya" V 1-SG3-SP VFIN { he } PR:na 1-SG2-OBJ OBJ { you } z
[nya] { shit } SV @FMAINVtr-OBJ>
"<ikinya>" "nya" V 9-SG-SP VFIN NO-SP-GLOSS COND:ki z [nya] { shit } SV
@-FMAINV-n
"<ilikunya>" "nya" V 4-PL-SP VFIN { they } PAST INFMARK z [nya] { rain }
SV @FMAINVintr
"<nya>" "nya" <Heur> N 9/10-SG { nya } @SUBJ
"<wanaokunya>" "nya" V 2-PL3-SP VFIN NO-SP-GLOSS PR:na 2-PL-SUB-REL { who
} INFMARK z [nya] { drop } SVO MONOSLB @FMAINVtr-OBJ>
"<wanaokunya>" "nya" V 2-PL3-SP VFIN NO-SP-GLOSS PR:na 2-PL-SUB-REL { who
} INFMARK z [nya] { drop } SVO MONOSLB @FMAINVtr-OBJ>
"<anye>" "nya" V SBJN 1-SG3-SP VFIN NO-SP-GLOSS z [nya] { shit } SV
@FMAINVtr+OBJ>
"<akinya>" "nya" V 1-SG3-SP VFIN { she } COND-IF FUT:ta { when } z [nya]
{ shit } SV @FMAINVtr+OBJ>
"<anye>" "nya" V SBJN 1-SG3-SP VFIN NO-SP-GLOSS z [nya] { shit } SV
@FMAINVtr+OBJ>
"<akinya>" "nya" V 1-SG3-SP VFIN { he } COND-IF PAST { when } z [nya] {
shit } SV @FMAINVtr+OBJ>
```

If we also want context, the process is more complicated but manageable. The result of (2) above will be analysed, disambiguated, and processed, so that it has the format as in (5). In order to save space, only part of result is displayed.

(5)

```
Kukojoa {kooja_V} na {na_CC} kunya {nya_V} kitandani {kitanda_N} : Mtoto
{mtoto_N} anapokojoa {kooja_V} kitandani {kitanda_N} baada_ya
```

{baada_ya_PREP} miaka {mwaka_N} miwili {wili_NUM} isionekane {onekana_V}
kama {kama_ADV} jambo {jambo_N} la_ajabu {la_ajabu_ADJ} .
Maaluni {maaluni_N} hawa {hawa_PRON} hata {hata_ADV} huko {huko_ADV}
kwenye {kwenye_PREP} mwezi {mwezi_N} watakakokwenda {kwenda_V} siku
{siku_N} wakikuta {kuta_V} mavi {mavi_N} watasema {sema_V} sisi
{sisi_PRON} ndiyo {ndiyo_ADV} tuliokunya {nya_V} huko {huko_ADV} .
Kivunya {Kivunya_Heur} alisema {sema_V} kabla {kabla_ADV} hawajamvamia
{vamia_V} , watu {mtu_N} hao {hao_PRON} walidhibitiwa {dhibitiwa_V} na
{na_CC} kupata {pata_V} kipigo {kipigo_N} kutoka_kwa {toka_kwa_PREP}
walinzi {mlinzi_N} wa {wa_GEN-CON} Bw {Bw_N} .
Alisema {sema_V} hayo {hayo_PRON} akikumbuka {kumbuka_V} mambo {jambo_N}
ya {ya_GEN-CON} utoto {utoto_N} alipokuwa {wa_V} akirudi {rudi_V} usiku
{usiku_ADV} siku {siku_N} za {za_GEN-CON} Jumamosi {jumamosi_N} alipoanza
{anza_V} mapenzi {penzi_N} na {na_CC} Jonas {Jonas_Heur} mpaka
{mpaka_CONJ} alipomwadhibu {adhibu_V} kwa {kwa_PREP} kumfina {fina_N}
na {na_CC} kisu {kisu_N} kwenye {kwenye_PREP} mapaja {paja_N} .
Mkurugenzi {mkurugenzi_N} huyo {huyo_PRON} aliwataka {taka_V} wananchi
{mwananchi_N} kujihadhari {jihadhari_V} na {na_PREP} ugonjwa {ugonjwa_N}
huo {huo_PRON} na {na_CC} kuacha {acha_V} kunya {nya_V} juisi {juisi_N}
mitaani {mtaa_N} .
Mama {mama_N} atuhumiwa {tuhumiwa_V} kumfina {fina_V} kwa {kwa_PREP}
ncha {ncha_N} ya {ya_GEN-CON} kisu {kisu_N} ukeni {uke_N} bintiye
{binti_N} mdogo {mdogo_N} .
Aliwataja {taja_V} watuhumiwa {mtuhumiwa_N} hao {hao_PRON} kuwa
{kuwa_CONJ} ni {ni_V} Patrick {Patrick_Heur} Mahinya {Mahinya_Heur} (35
{35_NUM}) na {na_AG-PART} Mapesa {pesa_N} Mkwavi {Mkwavi_Heur} (17
{17_NUM}) .
Kunya {nya_V} ni {ni_V} aula {la_V} zaidi {zaidi_AD-ADJ} kwa {kwa_PREP}
kukatazwa {katazwa_V} mwingine {ingine_ADJ} hakihitajii {hitajia_V} .

We see that there are words with the stem {nya_V} but also other stems ending with the characters *nya*.

When we use an advanced search system, we can retrieve only those sentences, which contain the stem *nya* (6).

(6)

Haya, ikawa vua inakunya {nya_V} kunya ngia ga uko kati.

Bali aaah kwa vua inakunya {nya_V} uzuri.

Akenda kunya {nya_V} haja chumbani kule akaja kunya {nya_V} ukumbini
akenda kunya {nya_V} barazani.

Akaja kunya {nya_V} ukumbini, akenda kunya {nya_V} barazani, jamaa wale.

Hapo ilianza kunya {nya_V} mvua asubuhi, mvua iliiliendelea mpaka
ikafika saa sita.

Anye {nya_V}, akesha kunya {nya_V} zile_zile kata -- kanapakusiwa,
usubuhi kazi iwapo.

Vyama vingine vinavyofikiriwa kushirikiana na chama cha leba katika
serikali ni MERETS kilichoshinda viti 9, Yisrael-ba alinya {nya_V} chama
ambacho ngome kuu ni wahamiaji ambacho kina viti 7.

Kukojoa na kunya {nya_V} kitandani: Mtoto anapokojoa kitandani baada_ya miaka miwili isionekane kama jambo la_ajabu.

Maaluni hawa hata huko kwenye mwezi watakakokwenda siku wakikuta mavi watasema sisi ndiyo tuliokunya {nya_V} huko.

Mkurugenzi huyo aliwataka wananchi kujihadhari na ugonjwa huo na kuacha kunya {nya_V} juisi mitaani.

Kunya {nya_V} ni aula zaidi kwa kukatazwa mwingine hakihitaji.

Bila kusita niliwapa wazi kuwa wao ni wezi na mafisadi wanaokunya {nya_V} kwenye ofisi zetu huku wakijiona wajanja.

Kunya {nya_V} anye {nya_V} kuku, akinya {nya_V} bata kaharisha.

We managed to retrieve the occurrences of the monosyllabic verb *nya* from the corpus of 25 million words.

2.2 The verb *ku-cha* (to rise, to fear)

We make another test with the verb *ku-cha* (to rise, to fear). This task is much more difficult, because the string *cha* is also a genitive connector in class 7 (e.g. *kitabu cha mtoto*, child's book). When our key word in search is '*cha*', we get 123995 occurrences, that is, vastly more than we can conveniently analyse.

We take a look at the result, so that we can see, how we could safely reduce the result (7).

(7)

Chuo cha Taifa cha Utalii Nchini kimepanga ku huo kinachotumiwa kiko kimoja cha Forodhani ambacho hivi sas hatua zinazochukuliwa na chuo cha Taifa cha Utalii katika ku zochukuliwa na chuo cha Taifa cha Utalii katika kuingia kati a na ukweli kuwa hakuna kikao cha Kamati ya Utendaji ya klab ilisema kama kweli kuna kikao cha Kamati ya Utendaji iliyoku Mtibwa ilitaka kumuacha Costa ikiwa tu Simba itaku Kipigo chake cha kwanza ni kuiteka Dar es s Chama cha upinzani cha CUF kimeanza Chama cha upinzani cha CUF kimeanza mikakati ya n Licha ya ajenda hiyo ya maombole e umuhimu wa kuipa CCM kipigo cha aibu katika uchaguzi mkuu jeruhiwa vibaya na wengi wakaacha makazi yao na kutimkia huk ya maandamano katika kipindi cha takribani wiki moja tu. a yaliko Makao Makuu ya chama cha upinzani cha TLP, chama am o Makuu ya chama cha upinzani cha TLP, chama ambacho kimekuw aka mgambo waliompiga mpiga picha waadhibiwe vikali. Chama cha Wapigapicha za Habari Nchi Chama cha Wapigapicha za Habari Nchini, PPAT, ki rifa hiyo akasema kuwa mpigapicha huyo alikuwa katika majuku inondoni kukifuatilia kikundi cha mgambo hao kilichofanya sh ya shambulio dhidi ya mpiga picha huyo na kukichukulia hatua lipokuwa katika eneo la kituo cha mabasi cha Ubungo ambapo l tika eneo la kituo cha mabasi cha Ubungo ambapo licha ya kuw ha mabasi cha Ubungo ambapo licha ya kuwaonyesha kitambulish sha kitambulisho chake halali cha kazi kinachotolewa na Seri

bia za kunyanyasa wagonjwa kuacha mara moja.
ea maisha bora zaidi, kwani licha ya kupata mshahara kazini
ayo imefanya kazi kwa kipindi cha miaka mitano, imekusanya p
u, umehifadhiwa katika chumba cha kuhifadhia maiti katika ho
a kuongeza kuwa marehemu hakuacha ujumbe wowote na wala saba
toka kilabuni kupata kinywaji cha pombe za kienyeji.

We see that we can remove the string ' cha ' (space on both sides), because the only context, where the form could occur as a verb is imperative singular, which is very unlikely. When we do this, we get the result as in (8).

(8)

aaluma mkoani, hivyo kwa kuwaacha walimu hao kutoshiriki zoe
ika ajali hiyo ambapo pia imeacha yatima na pia kuwaacha wen
ia imeacha yatima na pia kuwaacha wengine wakiwa na vilema v
ngeta Kasika (54), ambaye ameacha watoto wanane ambao waliku
li hiyo ambapo Rhoda naye aliacha watoto watatu.

Marehemu wameacha watoto ambao walitegemea k
Ajali hii mbaya pia iliacha Watanzania kila mmoja akiw
Jeshi Stars, iliyosaniwa na kocha wake mkuu, Martin Kemwaga

Licha ya Bunge kutangaza kumpa m
wanalipwa mafao yao mapema licha ya kasoro iliyojitokeza," a

Kwa mujibu wa Bw. Nyundo, licha ya mchango wa Manispaa, Se
Aibu kuikacha Simba - Asili.

ji takribani watano enzi za kocha wao Mkongo, Raoul Shungu.
walivyosema kuwa Yanga imewakacha ni sahihi kabisa.

Kocha wa timu hiyo, William Kima
Mkijificha kwenye makaburi, nani ataw
wakienda huko, lazima watajificha tu wakiogopa kukamatwa," a
amewaasa viongozi wa dini kuacha kutoa kauli zinazoweza kuh
inyonga na katika ujumbe alioacha amedai kuwa amejimaliza ku
asema kabla ya kujinyonga ameacha ujumbe unaosema kwamba ame
Nimeacha mke na mtoto mmoja.

amewataka wakazi wa Sinza kuacha dharau kwa kuwa gonjwa hil
Licha ya kumuua mfanyabiashara h
Mpigapicha Kassim Makongoro, 48, anay
na angependa Kassim ampigie picha mbili za kawaida toka kwen
Alitaka picha anayoingia ukumbini na ali
na siku zinakwenda na haoni picha zake zikiletwa, alinieleza
wa makubalianio ya kumpigia picha za kawaida toka kwenye mka
ni wamewataka viongozi wao kuacha kuzungumza uongo katika vy
ichezo Taifa, BMT, kutaka makocha wote akiwemo yeye kuwasili
eye kuwasilisha CV zake za ukocha IFCA.

ibuni BMT ilitangaza kuwa makocha wote Tanzania nzima wajian
a CV zao katika vyama vya makocha vya wilaya ili waweze kuwa
Wakati huo huo, kocha Boniface Mkwasa tayari ame
la aliyasema hayo jana kuwa kocha huyo amepeleka vyeti juzi
John Pombe Magufuli leo amewaacha wabunge midomo wazi pale a

Akasema mwanamama huyo licha ya kubembelezwa na mumewe
aga michezo hiyo, na hivyo kuacha kibarua kwa Mtagwa ambaye
wa kupata medali, na hivyo kuacha kazi kwa Mtagwa na Aziz Mw
Licha ya kumuua mfanyabiashara h

Here we can see, how we could do further pruning. We remove such strings, which cannot be forms of the verb 'cha' (9).