

**TEKNOLOJIA YA LUGHA KATIKA UTAFITI WA KISWAHILI: KIFANI CHA
MRADI WA SALAMA**

GIKAMBI, HEZEKIEL PETER

**TASNIFU HII IMETOLEWA ILI KUTOSHELEZA BAADHI YA MAHITAJI
YA SHAHADA YA UZAMILI KATIKA CHUO KIKUU CHA NAIROBI**

OKTOBA, 2015

UNGAMO

**NAKIRI KUWA TASNIFU HII NI KAZI YANGU ASILIA NA HAIJAWAHI
KUWASILISHWA KWA UTAHINI KATIKA CHUO KIKUU KINGINE CHOCHOTE**

GIKAMBI, HEZEKIEL PETER

TAREHE

**TASNIFU HII IMETOLEWA KWA IDHINI YETU KAMA WASIMAMIZI
WALIOTEULIWA WA CHUO KIKUU CHA NAIROBI**

DKT. OMBOGA ZAJA

TAREHE

DKT. IRIBE MWANGI

TAREHE

TABARUKU

Naitabaruku kazi hii kwanza kwa mke wangu nimpendaye Bi. Kawira Z Gikambi na wanetu wependwa binti zetu Furaha Gikambi, Faraja Gikambi na ghulamu Fadhili Bin Gikambi.

Nyinyi ndio motisha wa bidii yangu

Shukran

Kwa maombi yenu na uvumilivu wenu wakati nikikisha kwenye kompyuta.

Pili kwa wataalamu na wanafunzi wote wa Teknolojia ya Lugha ya Kiswahili wa sasa na watakaonifuatia. Isimu Kongoo (*Corpus Linguistics*) katika Kiswahili ni jua ambalo tayari

limechomoza!

Tulikaribishe na kulichangamkia.

SHUKURANI

Ningependa kutanguliza shukran zangu kwa Mwenyezi Mungu mwingi wa rehema aliyenipa afya njema hadi nikaweza kumudu kusoma shahada hii katika hali ngumu na changamoto za kila aina. Katika mazingira niliyokamilisha kazi hii, ilihitaji mkono wako. Wewe ni mwaminifu milele.

Pili, napenda kutoa shukrani zangu za dhati kwa wasimamizi wangu Dkt Zaja Omboga na Dkt Iribe Mwangi kwa kuniongoza na kunihimiza niuvalie njuga utafiti wa eneo changa kabisa la teknolojia ya lugha ambalo lilikuwa kwangu tayari limegeuka kutoka ari na kuwa kama uraibu. Kwa kunipa moyo, kunielewa na hata kunivumilia nilipokuwa nang'ang'ana na changamoto za kikazi kando na masomo yangu. Imani yenu katika uwezo wangu ilinifanya nikeshe ili niweze kupata muda wa kutafitia kazi hii. Nimejifunza mengi kutoka kwenu ila sina cha kuwalipa zaidi ya kuwaombea kwa Mwenyezi Mungu awajalie maisha mema. Asanteni sana wasimamizi wangu na walimu wangu wapendwa. Ninaamini Mungu, kama mlivyonishauri, sitakomea hapa.

Tatu, natoa shukrani za dhati kwa mke wangu nimpendaye Kawira Gikambi na wanetu wapendwa Furaha binti Gikambi, Faraja Binti Gikambi na Fadhili Bin Gikambi. Mlikubali kunikosa mara kwa mara mkiwa na matumaini kwamba muda utakapotimia nitamaliza masomo yangu na kisha kupata muda wa 'familia' wa kutosha pamoja nanyi. Nawashukuru nyote kwa moyo wa uvumilivu na upendo mkubwa.

Nne, pia nawashukuru wazazi wangu walionilea, kunishauri na kunisomesha. Mamangu Jeniffer Thirindi Peter kwa imani yako kuu katika uwezo wangu. Hili daima lilinisukuma mbele. Mungu akupe maisha marefu. Kwa babangu Marehemu Mzee Peter M'Ituiri M'Nabea Uliponiambia nisome 'hadi nipande ndege'... nilijua haukuwa mzaha kwako. Ombi hilo tayari Mungu kalitimiza, na huwa nalikumbuka ninapotembea ulimwengu mzima kwa sababu ya lugha hii ya Kiswahili ila hukubahatika kuishi kushuhudia ombi lako likitimia. Najua mahali ulipo unatabasamu.

Tano, nawashukuru walimu wangu katika Chuo Kikuu cha Nairobi, Idara ya Kiswahili walionifunza katika viwango tofauti na kunipa msingi bora wakiongozwa na mwenyekiti wa Idara ya Kiswahili Prof. John H Habwe ambaye kando na kunifundisha alinifaa kwa ushauri, maamuzi na maelekezi ya kipekee. Wahadhiri wa masomo mbalimbali walioniandaa Prof. Kithaka wa Mberia, Prof. Kyalo Wadi Wamitila, Prof. Mohammed Hassan Abdulaziz, Prof.

Rayya Timmamy, Prof. Kineene wa Mutiso, Dkt. Zaja Omboga na Dkt. Iribe Mwangi (waliokuja kuwa wasimamizi wangu katika kazi hii), Dkt Ayub Mukhwana, Dkt. Swaleh Amir, Mwalimu Basilio G. Mungania, Dkt. Alfred Buregeya, Dkt. Evans Mbuthia na Bi. Mary Ndungu siwezi kuwashukuru vya kutosha kwa kunipa msingi huu, wengi wenu kutoka shahada ya kwanza, ninapojiandaa kuingia katika ulimwengu wa usomi. Mungu awabariki sana.

Tano, kwa wenzangu katika darasa hili la Uzamili - 2015, (Watoto wa Babu) tulikuwa na tukaendelea kuwa kama watoto kumi wa mama mmoja; Stephen Machira, Biasha Salim, Francisco Gathara, Henry Indindi, Rose Kamwanza, Rita Magara, Makhulo Evans, Rose Dama na Timothy Otuchi. Sijawahi kuwa katika darasa kama hili tangu nianze masomo yangu. Mungu awabariki na udumu undugu wetu.

Sita, kwa Ndugu Prof. Mstaafu Arvi Hurskainen wa Chuo Kikuu cha Helsinki, Finland, mwasisi wa mradi wa SALAMA tulioutafitia. Bila wewe utafiti huu ungekuwa haupo kabisa. Nakushukuru kwa mawasiliano, maelekezi na majibu ya kina ya hojaji zangu. Ulitutangulia na kutuonyesha njia. Mungu akulinde. Siwezi kumsahau Dkt. Suleiman S. Sewangi – Katibu Mtendaji wa BAKITA- Tanzania na mwanafunzi wa zamani Prof Hurskainen, kwa majibu na ushauri wako kuhusu njia hii uliyokwishapitia.

Pia, shukran za kipekee ziwaendee Prof Assibi Amidu –Chuo cha NTNU-Trodheim-Norway kwa vitabu vyako ulivyoniletea kwa posta na kwa undugu wetu, kwa Prof. Clarissa Vierke-Chuo Kikuu cha Bayreuth Ujerumani kwa kunialika na kunikaribisha na kunipa fursa ya kusimama mbele za wasomi watukuka kutoka kote ulimwenguni nikiwasilisha makala katika makongamano ya Bayreuth na kwa Prof. Dr. Rose Marie Beck na Dkt Irene Brunnotti kwa kunialika Chuo Kikuu cha Leipzig-Ujerumani kunifungua macho kimasomo. Asanteni kwa ushirikiano na ukarimu wenu. Mwisho, ni kwa wenzangu katika Kampuni ya Usambazaji Habari ya Nation Media Group, siwezi kuwataja nyote ila naridhia mwongozo na msaada mahsusni wa Tom Mshindi, kwa kunipa fursa ya kusoma nikiwa kazini. Prof. Ken Walibora, kakangu na msimamizi wangu kazini ambaye daima hunipa ushauri adimu ninapiga hatua za kuelekea alikofika zamani, asante kwa motisha na uelekezi bora. Kwa wenzangu katika Taifa Leo, Swahilihub, NTV, QFM, QTV na magazetini, pamoja tulielea, hakuna fursa ya kuzama! Mungu yupo. Kwa wengine wote mlionifaa kwa hili au lile, Mungu awajazie baraka tele. Shukran jazilan.

YALIYOMO

Ungamo.....	(ii)
Tabaruku	(iii)
Shukran	(iv)
Yaliyomo.....	(vi)
Iksiri	(x)
Ufafanuzi wa dhana muhimu	(xi)

SURA YA KWANZA

UTANGULIZI WA UTAFITI

1.0 Utangulizi.....	1
1.1 Tatizo la Utafiti	4
1.2 Malengo ya utafiti	7
1.3 Maswali ya Utafiti	8
1.4 Sababu za Kuchagua mada	8
1.5 Upeo na mipaka ya Utafiti	8
1.6 Yaliyoandikwa kuhusu mada hii/Udurusu wa Machapisho	11
1.7 Msingi wa Nadharia ya Utafiti.....	16
1.7.1 Nadharia ya Msambao Wa Ugunduzi	17
1.7.2 Nadharia ya Isimu Kongoo au Mtazamo wa Kikongoo.....	17
1.7.2.1 Nguzo za Nadharia ya Isimu Kongoo	20
1.7.2.2 Udhaifu wa Nadharia ya Isimu Kongoo	24
1.8 Nadharia tete za Utafiti/Haipotesia za utafiti	24
1.9 Mbinu za Utafiti.....	25
1.9.1 Ukusanyaji wa data.....	25
1.6.1.1 Aina za data.....	25
1.9.2 Uchanganuzi wa data	28
1.9.3 Uwasilishaji wa data	28

SURA YA PILI

PROGRAMU ZILIZOUJENGA MRADI WA SALAMA

2.0 Utangulizi.....	29
2.1 Historia fupi ya Mradi wa SALAMA	29
2.2 Kichanganuzi cha Kimofolojia- SWATWOL	31
2.2.1 Sifa za kimuundo za SWATWOL	33
2.2.2 Msamiati uliofafanuliwa katika mfumo wa tabaka-mbili	34
2.2.3 Kanuni za Tabaka-mbili	35
2.3 Kitatuzi cha utata wa maana na kichambuzi msingi cha Sintaksi- SWACGP.....	36
2.4 Kitatuzi-Kibahatishi cha Utata wa Matini za Kiswahili cha SWA–GUESS.....	38
2.4.1 Kanuni za kutambua maana halisi kwa Ubahatishaji wa kujaribujaribu	39
2.4.2 Kubahatisha kwa Misingi ya Uwezekano.....	39
2.5 Kichanganuzi cha kina cha kisintaksia cha SWAFGP	40
2.6 Kitafsiri-Oto cha Kiswahili kwenda InterLingua na Mifumo mingine	41
2.7.1 Kanuni za kubadilisha uainishaji wa maneno	41
2.7.2 Kanuni za kupanga maneno-upya	42
2.8 Programu ya Uwakifishaji ya – SWASENT	42
2.9 WLIST- Kuorodheshaji cha maneno.....	42
2.10 Programu ya Kuunda na kutathmini Kamusi ya SALAMA-DC	43

2.11 Hitimisho.....	44
---------------------	----

SURA YA TATU

3.0 KONGOO-MATINI KATIKA TEKNOLOJIA YA LUGHA

3.1 Utangulizi.....	46
3.2 Maana ya Kongoo/Kopasi.....	46
3.3 Ukubwa wa Kongoo.....	47
3.4 Aina Za Kongoo.....	49
3.4.1. Utanzu wa matini	49
3.4.2. Aina ya Data katika kongoo.....	49
3.4.3. Aina ya Matini katika kongoo.....	51
3.4.4 Lengo la Mkao wa Kongoo.....	51
3.4.5 Aina ya Matumizi ya Kongoo.....	52
3.5. Sifa kuu na muhimu za kongoo.....	53
3.5.1 Ukubwa.....	53
3.5.2. Ubora wa kongoo (uasilia).....	54
3.5.3 Uwakilishaji.....	54
3.5.4 Usawa.....	54
3.5.5 Urahisi.....	54
3.5.6 Upatikanaji	54
3.5.7 Uthibitishikaji.....	55
3.5.8 Mpangilio wa habari kwenye kongoo.....	55
3.5.9 Uongezekaji.....	55
3.6 Kongoo Ya Kiswahili ya Helsinki- SALAMA (HCS).....	55
3.6.1 Vyanzo vya Kongoo ya Helsinki Corpus of Swahili (HCS).....	56
3.6.2 Historia ya Kongoo ya Kiswahili ya Helsinki (HCS).....	58
3.6.3 Masharti ya kuitumia kongoo ya Kiswahili ya Helsinki (HCS).....	59
3.7 Nafasi ya Kongoo ya HCS katika Mradi wa SALAMA.....	60
3.7 Manufaa ya kongoo katika utafiti.....	61
3.7.1 Matumizi mahsusi ya kongoo katika utafiti wa lugha.....	61
3.8 Hitimisho.....	63

SURA YA NNE

UCHAMBUZI NA UFAFANUZI WA DATA-TEKNOLOJIA YA LUGHA KATIKA UTAFITI

4.0. Utangulizi.....	64
4.1. Uwasilishaji na uchanganuzi wa data	64
4.2 Teknolojia ya lugha katika utafiti wa Mofolojia ya Kiswahili.....	65
4.2.1 Uchanganuzi wa kimofolojia kutumia SWATWOL.....	66
4.2.2 Kushirikisha habari za Kisemantiki katika uainishaji wa kimofolojia.....	68
4.2.3 Kushirikisha uchambuzi wa kisintaksia katika uainishaji wa kimofolojia	69
4.2.4 Kuelezea tungo-funge kimofolojia katika tafsiri.....	70
4.2.5 Kudhibiti Mpangilio wa maneno	71
4.2.6 Uchanganuzi wa data	72
4.3 Teknolojia ya lugha katika utafiti wa Semantiki ya Kiswahili	73
4.3.1 Utata wa maana na Utatuzi wake katika Maneno ya Kiswahili.....	74
4.3.2 Uondoaji wa utata wa kisemantiki kutumia tafsiri-mashine na programu mseto.....	77
4.3.3 Kupata viwango vya ufasaha katika utatuzi wa utata wa kisemantiki	85
4.4. Teknolojia ya lugha katika utafiti wa Sintaksia ya Kiswahili.....	85
4.4.1 Uchanganuzi wa data kuhusu utafiti wa sintaksia kikompyuta	89

4.5.4.5	Teknolojia ya Lugha katika Uundaji na Utathmini wa Kamusi za Kiswahili.....	91
4.5.1	Uundaji wa vidahizo vya kamusi	92
4.5.1.1	Uundaji wa vidahizo vya neno moja.....	92
4.5.1.2	Uundaji wa vidahizo vya tungo-funge.....	93
4.5.1.3	Uundaji wa vidahizo vya kipolisemia.....	94
4.5.1.4	Uundaji wa vidahizo vya kihomonimu	94
4.5.1.5	Urejeleano wa vidahizo kwenye kamusi.....	95
4.5.2	Uchopoaji wa mifano ya kufafanulia maana ya vidahizo katika kamusi.....	96
4.5.2.1	Mifano kwa kufupisha muktadha wa matumizi.....	97
4.5.2.2	Kuchagua mifano	97
4.5.2.3	Uteuzi -nasibu wa mifano uliodhibitiwa.....	97
4.5.2.4	Uteuzi kwa misingi ya muktadha wa idadi ya utokeaji	98
4.5.3	Kuunganisha vijisehemu ili kuunda kamusi	98
4.5.4	Kutafsiriwa kwa mifano katika kamusi iliyoundwa	100
4.5.5	Utathmini wa SALAMA-DC kutumia kongoo-matini tatu za Kiswahili	102
4.5.6	Jinsi ya kutumia kamusi hii.....	103
4.5.7	Matumizi ya SALAMA katika tathmini ya Kamusi tano za Kiswahili	103
4.5.8	Uchanganuzi wa data za uundaji na utathmini wa kamusi kikompyuta	109
4.6	Teknolojia ya Lugha, Tafsiri-Mashine na Utafiti katika Kiswahili.....	110
4.6.0.	Utangulizi.....	110
4.6.1.	Tafsiri-mashine na vivumishi katika Kiswahili	111
4.6.2.	Tafsiri-mashine ya Maswali ya Kiswahili kwa Kiingereza	111
4.6.3	Tafsiri-mashine ya Biblia ya Kiswahili kwa Kiganda	116
4.6.4.	Tafsiri-mashine ya Nomino za pekee kwenda Kiingereza.....	119
4.6.5.	Tafsiri-mashine ya Majina ya Watu na siku za Wiki.....	119
4.6.6.	Tafsiri-mashine ya vitenzi na Minyambuliko ya Kiswahili kwenda Kiingereza	119
4.6.7.	Tafsiri-mashine na Virejeshi vya Kiswahili kwa Kiingereza.....	120
4.6.8	Tafsiri-mashine na tungo-funge za Kiswahili kwa Kiingereza.....	120
4.6.9	Uchanganuzi wa Data	121
4.7	Teknolojia ya lugha katika Utafiti wa Fasihi ya Kiswahili	121
4.8	Teknolojia ya Lugha na Ufundishaji wa Kiswahili- CALL.....	122
4.9	Teknolojia ya lugha katika Utafiti wa Isimu Jamii- Mabadiliko ya Kiswahili	126
4.10	Teknolojia ya Lugha Utafiti wa Istilahi za Sajili mahsusi	127
4.11	Teknolojia ya lugha katika Ukaguaji tahajia na sarufi Kiswahili.....	127
4.12	Kongoo katika Teknolojia ya lugha na Utafiti wa Kiswahili.....	127
4.13	Hitimisho.....	129

SURA YA TANO

5.0	Utangulizi.....	130
5.1	Muhtasari wa Utafiti	130
5.2	Tathmini ya Maswali ya Utafiti na Upimaji wa Haipotesia.....	132
5.3	Maoni ya Kijumla/Mchango wetu	134
5.3.1	Ufaafu wa nadharia za utafiti.....	135
5.3.2.	Changamoto za utafiti	135
5.4	Mapendekezo ya utafiti.....	135
5.5	Hitimisho.....	139

MAREJELEO	141
------------------------	------------

Viambatisho

Kiambatisho 1: Ukurasa wa kwanza wa tovuti ya SALAMA (Swahili Language Manager).....	147
Kiambatisho 2: Jedwali linaloonyesha hatua tofauti na majukumu ya mazingira ya SALAMA...	148
Kiambatisho 3: Matumizi ya programu huria katika utafiti wa Kiswahili kwa mitazamo ya kikongoo (Mfano wa Anticon3 na upatanisho wa sarufi wa maneno kwenye matini.....)	149
Kiambatisho 4: Sampuli ya maswali ya hojaji iliyopelekewa mwasisi wa SALAMA Prof. Arvi Hurskainen	150
Kiambatisho 5: Sampuli ya maswali ya hojaji iliyopelekewa kwa wataalamu walioshiriki katika utafiti wa mradi wa SALAMA Dkt Suleiman Sewangi na Dkt Wanjiku Ng'ang'a	152
Kiambatisho 6: Faharasa ya alama za kimofolojia zilizotumiwa katika Tafsiri Mashine .	154

IKISIRI

Utafiti wa lugha kuegemea mitazamo ya kikompyuta umetamalaki katika lugha zilizoendelea kidijitali kama Kiingereza na nyinginezo ila dhana hii haijakita mizizi kabisa katika utafiti wa lugha za Kiafrika kama Kiswahili. Utafiti huu umeangazia suala la matumizi ya teknolojia ya lugha katika utafiti wa Kiswahili kwa kuchunguza na kutathmini mbinu mbali mbali zilizotumika katika mradi wa SALAMA (Swahili Language Manager), ulioasisiwa na Arvi Hurskainen profesa wa taaluma za lugha za Kiafrika katika Chuo Kikuu cha Helsinki nchini Finland kutoka 1985- 2015 ili kutambua ukweli wa hali ya utafiti wa Kiswahili katika jukwaa la teknolojia ya lugha. Kwa kutumia mazingira ya SALAMA, ambayo yalishughulikia Kiswahili na kuchunguza nyanja zake mbali kwa kutumia programu zilizojengwa kwenye misingi ya kanuni za lugha na za kikompyuta, utafiti huu ulilenga kubainisha mbinu mbalimbali zilizohusika katika utafiti wa kikompyuta wa lugha ya Kiswahili katika mradi huu na kuzichunguza mbinu hizo zilizotumika kufanyia utafiti na kuchanganulia vipengele tofauti tofauti vya lugha hii, na iwapo zinaweza kuwa zimechangia katika kukuza na kuendeleza utafiti wa Kiswahili kama lugha ya Kiafrika inayoweza kutafitiwa kwa kutumia Teknolojia ya Lugha au Uhandisi Lugha. Katika kufanya hivi, Utafiti umelenga kuonyesha umuhimu, uamilifu na utumiwaji wa kongoo au kopasi katika Teknolojia ya Lugha na jinsi kanzi hii inavyoweza kuboresha utafiti katika lugha ya Kiswahili kikompyuta. Nadharia ya Msambao wa Ugunduzi (Diffusion of Innovation Theory) iliyoasisiwa na Everett Rogers (2003) pamoja na Nadharia ya Isimu Kongoo au Mtazamo wa Kikongoo zimetumika katika kujaribu kueleza jinsi teknolojia ya lugha ni mkabala ulio na manufaa zaidi kiutafiti kuliko mbinu za kimapokeo. Kupitia kwa data kutoka kwa hojaji na mifano kutoka kwa ripoti mbali mbali na kazi za kitaaluma zinazohusu SALAMA, utafiti umebaini kuwa teknolojia ya lugha inachangia katika utafiti ya kiujarabati zaidi, ina uwezo wa kuichakata data kubwa zaidi kikompyuta kuliko uwezo wa binadamu, inampa mtafiti mnepo na uhuru zaidi, inahifadhi lugha kikompyuta na inaweza kushughulikia lugha ya Kiswahili kiutafiti katika maeneo yote kutoka uundaji na utathmini wa kamusi, tafsiri-mashine, uchanganuzi wa kimofolojia, kisintaksia, kisemantiki, ufundishaji lugha na miongoni mwa masuala mengine.

UFAFANUZI WA DHANA MUHIMU NA VIFUPISHO

ACALAN- (*African Academy of Languages*) - Taasisi ya lugha ambayo ni kitengo cha Muungano wa Afrika (AU) iliyo na makao yake makuu jijini Bamako, Mali.

AU- African Union (Muungano wa Afrika)

CALL- Computer Assisted Language Learning (Njia za kujifunza lugha kutumia kompyuta).

COCA- (*The Corpus of Contemporary American English*) yaani Kongoo ya Kingereza cha Kisasa cha Kimarekani.

*Compatibility-*Ulinganaji/Ufaafu wa ugunduzi au teknolojia mpya husika.

*Complexity-*Uwezekano wa kimatumizi ya teknolojia mpya au uvumbuzi

Computational Linguistics- Isimu-Kompyuta (Tawi la isimu linalochunguza lugha kutumia mbinu za kikompyuta).

Computer Concordancing Programs- Programu za kompyuta za kuunda upatanisho wa lugha

Corpus Linguistics – Isimu Kongoo (Tawi la Isimu Kompyuta ambalo huhusika na uchunguzi wa matini ya lugha halisi katika matumizi yake kwenye mkusanyiko wa hifadhi za kikompyuta uitwao kongoo.

DAHE- Dar es Salaam- Helsinki- Mradi wa utafiti uliohusu kukusanya msamiati wa Kiswahili uliofanywa kwa Ushirikiano wa Chuo kikuu cha Helsinki na Dar es Salaam.

HSC- HELSINKI CORPUS OF SWAHILI - (Kongoo ya maneno milioni 12.5 ambayo ilitumika katika mradi wa SALAMA katika Chuo cha Helsinki Finland na inapatikana katika hifadhi ya Lugha ya Finland (Language Bank of Finland) kwenye tovuti www.csc.fi. Watafiti na taasisi za utafiti hutafuta ruhusa ili kuitumia kwa kujaza fomu ya maombi katika kiungo hiki:

<http://www.csc.fi/english/customers/university/useraccounts/languagebank>

ICALL- (*Intelligent Computer Assisted Language Learning*) - Mfumo wa kufundisha lugha

kutumia kompyuta na teknolojia ya lugha ambao ni shirikishi na wa kimajibizano uliotumika katika *SWALEARN*.

Intuitive Approach -Mtazamo wa kiung'amuzi – Mbinu ya kutafitia lugha kutumia mifano na kanuni kutoka kichwani na kufikiria kuwa ni sahihi kwa ndivyo ilivyo. Mtazamo huu ni kinyume na Mtazamo wa Kiisimu-Kongoo ambao huonyesha ushahidi wa lugha kimatumizi kwa kikompyuta.

KWIC -*Key Word in Context* (yaani kuweka neno kuu kwenye muktadha ufao ili uweze kutafuta machapisho maktabani na kadhalika kutumia kompyuta zenye uwezo mdogo sana

KVCBLC -*Kiswahili Vehicular Cross Border Language Commission*- Kamisheni ya Kiswahili kama Lugha ya Mawasiliano ya Nje ya Kanda Nzima.

Corpus – Kongoo au Kopasi - Huu ni mkusanyiko wa maneno au ni kanzi ya maneno ya Lugha yaliyokusanywa na kuhifadhiwa kama yameainishwa au katika hali yake halisi ya kimatumizi. (Katika Kongoo-iliyoalamishwa kila neno limeainishwa kiisimu).

LT- *Language Technology* (Teknolojia ya Lugha)- Uhandisi Lugha

Uhandisi Lugha- *Language Engineering* (Teknolojia ya Lugha)

NLP- *Natural Language Processing* (Uchakataji wa Lugha asilia wa kikompyuta)

Observability- Ushuhudiaji au Uambukizanaji

MT- Machine Translation (Tafisiri-Oto au Tafisiri-mashine)

MRD- Machine Readable Dictionaries-MRD- Kamusi zinazosomeka kikompyuta

TM- Translation Memory (Programu za kuhifadhi tafisiri iliyofanyika kwa kulinganisha maneno yaliyotafsiriwa)

Relative advantage - Manufaa/Faida/Istihaki rejelewa ya teknolojia mpya au ugunduzi

SALAMA- Swahili Language Manager (Mazingira ya kikompyuta ya kuchanganulia na kufanyia utafiti Lugha ya Kiswahili yenye programu anuwai zinazotumia Teknolojia ya Lugha. Mradi huu ulianzishwa na Prof. Arvi Hurskainen katika Chuo cha Helsinki Finland.).

SALAMA-DC- (*SALAMA Dictionary Compiler*) ni programu ya SALAMA ya kuunda na kutathmini kamusi ya Kiswahili kwa kutumia kongoo iliyoainishwa kama kuibuka na vidahizo na kutoa mifano.

SAWA CORPUS- Kongoo-pacha au Kongoo-Sambamba ya Kiswahili na Kingereza iliyotumika na wataalamu watatu de Pauw, de Shryver na waiganjo katika utafiti wa MT (tafsiri-mashine).

SOM- (*Self Organising Map*) - Programu inayosaidia katika kuondoa utata wa kisemantiki katika ufundishaji wa lugha. Ilishughulikiwa na Ng'ang'a 2005.

SWALEARN- Programu ya kufundishia lugha kwa kutumia uchanganuzi wa kimofolojia na uondoaji wa utata wa maana iliyojengwa na mfumo wa Xerox.

SWATWOL- *Swahili Two Level Morphology*- Programu ya kuchanganua Lugha ya Kiswahili kimofolojia kwa kutumia mbinu ya Mofolojia ya Tabaka mbili na pia kukagua tahajia na kuchambua lugha kisintaksia.

SWACGP- *Swahili Constraint Grammar Parser*-(Programu ya kompyuta inayotekeleza majukumu ya kutatua utata wa maana ya maneno ya Kiswahili na pia kuchambua Lugha kisintaksia

SWASENT- Programu ya SALAMA ya kupanga sentensi katika mstari mmoja na kutekeleza majukumu ya uakifishaji.

SWALEARN- Programu ya kufundisha Kiswahili ya SALAMA inayotumia misingi ya kongoo na teknolojia ya lugha kwa jumla katika misingi ya uainishaji lugha na uondoaji wa utata wa maana.

Under-Resources Languages- Lugha zenye hazina haba za kiutafiti

WLIS- Programu ya SALAMA ya kupanga matini za Kiswahili kwa mwegemeo wima kila neno katika mstari wake. Kuna aina nyingi za mbinu hii zikiwa mahsusi kwa kila jukumu.

SWAGUESS- Programu ya ALAMA iliyojengwa na Flex (mbinu za kampuni ya Unix)

inayoondoa utata wa kisemantiki wa matini kwa kubahatisha. Hushughulikia yaliyoachwa na *Constraint Grammar* - Sarufi Kikwazo.

SWAFDP- ni ufupi wa Kiingereza wa Swahili Functional Dependency Parser ambacho ni kichanganuzi cha kikompyuta chenye uwezo mkubwa kuliko **SWACGP** ambacho hufanya uchanganuzi wa kina wa kisintaksia na huweza hata hufafanua kwa kielelezo cha miti.

SWA-TO- InterLingua- Programu ya kugeuza sarufi na kutafsiri Kiswahili hadi Lugha unde ya InterLingua kwa kubadilisha SWAGUESS kidogo.

Trialability- Ujaribishikaji wa teknolojia mpya katika jamii kwa kiwango kidogo kwanza.

TUKI- Taasisi ya Uchunguzi wa Kiswahili

UDSM- University Of Dar Es Salaam (Chuo Kikuu cha Dar es Salaam)

WSD- Word Sense Disambiguation (Uondoaji wa Utata wa maana ya maneno kwa kutumia uainishaji wa kina zaidi wa kimofolojia na pia wa kisemantiki).

Xerox- Programu ya kushughulikia ujirudaji wa vipashio vya lugha kwa kuangalia uundaji wake.

Equality- Usawa

Simplicity- Urahisi

MAELEZO YA ISTILAHU ZA UAINISHAJI WA KONGOO

Genre of Text- Utanzu wa matini

Nature of Data- Aina ya data katika kongoo husika

Type of Text- Aina ya Matini katika kongoo

Purpose of Design- Lengo la Mkao wa Kongoo

Nature of Application - Aina ya Matumizi ya Kongoo ambayo kongoo inakusudiwa.

Quantity- Ukubwa wa Kongoo

Quality/Authenticity- Ubora wa kongoo au uhalisia wake

Representation- Uwakilishaji wa Nyanja tofauti za lugha katika kongoo

Equality -Usawa wa vijenzi vya kongoo kiasilimia

Simplicity- Urahisi- Kongoo haifai kuwa na uchangamano unaotatiza utafiti

Retrievability- Upatikanaji wa data kutoka kwa kongoo

Verifiability- Uthibitishikaji-Uwezo wa kuthibitisha data kutoka kwa kongoo.

Reference corpus- Kongoo Marejeleo

Spoken Corpus-Kongoo za Hotuba

General Corpus-Kongoo za jumla

Special corpus- Kongoo spesheli

Sub-language corpus- Kongoo ya sehemu ya lugha

Sample corpus- Kongoo sampuli

Literary corpus- Kongoo za kifasihi

Monitor corpus- Kongoo fuatili

Monolingual corpus -Kongoo ya lugha moja

Bilingual corpus- Kongoo ya lugha mbili

Parallel Corpus- Kongoo Sambamba

Multilingual corpus- Kongoo ya Lugha Nyingi

SURA YA KWANZA

1.0 UTANGULIZI

Teknolojia imekuwa pumzi ya ulimwengu katika karne ya 21 kwa kuwa inaendesha shughuli nyingi za binadamu. Teknolojia ya Lugha kwa upande mwingine ni mbinu inayozidi kuchukua nafasi muhimu katika utafiti wa lugha kote ulimwenguni na kufidia mapungufu ambayo yamekuwa yakishuhudiwa katika utafiti wa king'amuzi (intuitive research) au wa kimapokeo.

Maendeleo ya Teknolojia ya Lugha huwa yameshajiishwa na sera mbalimbali za lugha na teknolojia huwa inaimarishwa kwa lugha zenye majukumu muhimu ya kimawasiliano. Lugha za Kiafrika ambazo hazijapata fursa ya kuwepo katika jukwaa la kidijitali kwa muda mrefu kama baadhi ya lugha zenye asili ya Uropa zilivyo, pia zinatafitiwa na kuainishwa kutumia teknolojia ya lugha lau kwa viwango vidogo sana. Kiswahili ni mojawapo wa lugha za Kiafrika ambayo imedhihirisha uwepo mkubwa kwenye majukwaa haya ya kidijitali (De Pauw et al., 2006; De Pauw and de Schryver, 2008); Gikambi 2012, Muhirwe 2007). Katika majilio ya intaneti, lugha kadhaa za Kiafrika pia zimejikuta katika mazingira haya ya kidijitali-mtandaoni.

Kwa miaka mingi sasa wataalamu wengi wanajaribu kuangalia jinsi ya kuimarisha utafiti wa lugha za Kiafrika kwa kutumia teknolojia ya lugha katika kompyuta. Katika miongo miwili iliyopita, juhudi za kila aina zimefanywa katika teknolojia ya lugha za Kiafrika kwa kuvalia njuga juhudi za kuimarisha hazina ya nguvu-kazi, kongoo matini, miswada, makamusi, thesauri, na utafiti wa sarufi za lugha mbalimbali. Pamoja na hatua hizi hasa katika lugha ya Kiswahili suala ya usanifishaji linadhaniwa kama moja ya vikwazo vya kukabiliwa. Karne hii pia imeshuhudia hatua murua katika nyanja za ujanibishaji kama vile ujenzi wa programu za kikompyuta kama Kilinux (Kiswahili katika mfumo wa Linux) na kuwa na kitengo chake cha programu-huria ya Kiswahili yaani *Open Office* na pia tafsiri ya *MS Office 2003* na *Windows XP* kwa Kiswahili. Mradi wa Microsoft East Africa ulifuatia kwa kuwa na Kiolesura Fungasha cha Windows. Hili ni daraja la kuruhusu mtumiaji wa lugha ya Kiswahili kusanidi na kutumia programu za windows. (Mabeya, 2009). Kumekuwa na mwito kwa serikali za mataifa ya Afrika kufadhilia miradi ya kuweka programu huria za lugha zao kwenye majukwaa ya kidijitali. (Mabeya, 2009). Licha ya hatua hizi katika lugha ya Kiswahili haja hii bado ipo.

Utafiti wa lugha hizi kwa hivyo unalo hitaji kubwa la kuvumbua na kutumikiza programu imara za Teknolojia ya Lugha zinazoweza kuchakatia data za kiisimu. Data hii itakuwa na uwezo wa kutumiwa vizuri zaidi katika utafiti iwapo itakuwa imechanganuliwa kijarabati na kusomeka kikompyuta na pia kuhifadhiwa katika kompyuta (Hurskainen 2005). Data kubwa ya aina hiyo yaani kanzi ya maneno mengi tu yaliyokusanywa kwa pamoja jinsi yalivyotumika katika muktadha

wa mawasiliano kilugha ndiyo huitwa kongoo au kopasi. Kongoo huwa muhimu sana iwapo maneno yake ¹yamealamishwa au kuinishwa kwa kutolewa ufafanuzi wa kimsingi. Hata hivyo, kwa mujibu wa Sinclair (1976) kongoo ya maneno ya lugha katika hali yake halisi ya kimatumizi bila kualamishwa au kuainishwa inamfaa mtafiti yeyote ili aitumie atakavyo. Maoni haya yanaungwa mkono na Sewangi (2001).

Teknolojia ya Lugha haina budi kutumia kongoo/kopasi kama mali ghafi yake na pia kama chombo cha kufanyia utafiti katika nyanja mbalimbali za lugha. Mathalan, Kiswahili kimeshughulikiwa kikamilifu kwa kutumia kongoo iliyojengwa katika mazingira ya kikompyuta ya mradi wa SALAMA (Swahili Language Manager) ulioasisiswa na Prof. Arvi Hurskainen katika Chuo Kikuu cha Helsinki, nchini Finland. Kwa mujibu wa Hurskainen (1999), SALAMA ilianza kama programu ya kimsingi ya kuchanganua lugha kimofolojia na hakukuwa na mpango bainifu wa kuanzisha mradi huu. Hurskainen anasema kuwa mradi wa SALAMA umejengeka hatua kwa ya pili kulingana na mahitaji ya wakati huo.

Mradi wa SALAMA ulianzishwa mnamo mwaka wa 1985, na sasa unaongoza ulimwengu kama mradi unaoshirikisha maneno mengi zaidi ya Kiswahili kwa ukamilifu iwezekanavyo na kuhusisha nyanja kadhaa za lugha hiyo katika mfumo bainifu ya kuishughulikia lugha kikompyuta

(Hurskainen, 2004). Ni wazi kuwa lugha yoyote hubainika zaidi kiisimu kupitia kwa maneno yake kimaandishi. Maneno haya iwapo yanaalamishwa au yanaainishwa kwa ufasaha na kuchanganuliwa kutumia programu mbalimbali za kikompyuta, basi hayo ni mazingira mwafaka ya kujengea mfumo wa Teknolojia ya Lugha.

SALAMA inashirikisha programu mbalimbali za kikompyuta zilizoundwa kwa ushirikiano na wataalamu wa kikompyuta na wanaisimu. Programu hizo ni kama *SWATWOL- Swahili Two Level Morphology*- Programu ya kuchanganua Lugha ya Kiswahili kimofolojia- kutumia mbinu za Mofolojia ya Tabaka Mbili na kanuni za kimuundo ukomo, Hurskainen (2008).

....¹yamealamishwa (*Tagged corpus*)- Kualamisha kongoo ni kuainisha matini zake kwa kutumia programu mbalimbali za kikompyuta zinazoyapa maneno yake ufafanuzi wa kiisimu kama aina ya neno, ngeli yake, etimolojia, tafsiri ya lugha nyingine na kadhalika. (Hurskainen, 2004a) Haya yataangaziwa kwa kina katika sura ya pili na ya tatu.

Pia kuna programu ya SWACGP- ambayo ni akronimu ya Kiingereza cha Swahili Constraint Grammar Parser-(Programu ya kompyuta inayotekeleza majukumu ya kutatua utata wa maneno ya Kiswahili na pia kuchambua Lugha kisintaksia. *SWASENT* nayo ni programu ya SALAMA ya kupanga sentensi katika mstari mmoja na kutekeleza majukumu ya uakifishaji na *WLIST* Programu programu ya SALAMA ya kupanga matini za Kiswahili kwa mwegemeo wima kila neno katika mstari wake. Kuna aina nyingi za mbinu hii zikiwa mahsusi kwa kila jukumu. *SWAGUESS*-ni programu ya SALAMA iliyojengwa na Flex (mbinu za kampuni ya Unix) inayoondoa utata wa kisemantiki wa matini kwa kubahatisha. Hushughulikia yaliyoachwa na Sarufi Kikwazo (Constraint Grammar) kama tulivyodokezea katika faharasa ya istilahi, *SWAFDP*- ni ufupi wa *Swahili Functional Dependency Parser* (Kichambuzi cha Sarufi Amilifu Tegemezi ya Kiswahili) ambayo ni programu yenye uwezo mkubwa kuliko SWACGP ambayo hufanya uchanganuzi wa kina wa kisintaksia na huweza hata hufafanua kwa kielelezo cha miti, *SWA-TO- InterLingua*- Programu ya kugeuza sarufi na kutafsiri Kiswahili hadi lugha-unde ya InterLingua kwa kubadilisha SWAGUESS kidogo. Programu zinazounda mradi wa SALAMA zimetumiwa katika uundaji wa kamusi, kikagua tahajia (spell-checkers), na katika kutathmini kamusi mbalimbali miongoni mwa majukumu mengine ya kiutekelezi katika lugha hii. Pia kuna programu ya SWA-LEARN inayotumiwa katika shughuli za kufundishia lugha ya Kiswahili kwa Wageni na wanafunzi wa viwango mbalimbali. Programu hizi na mbinu nyingine za kikompyuta zimefafanuliwa kwa kina katika sura ya pili.

Hadi sasa programu zilizojenga mradi wa SALAMA zimetumika katika miradi mikubwa ya Kiswahili ulimwenguni kama vile kikagua tahajia cha programu za *Microsoft Office 2007 Suite*, **programu** za Kiswahili za Kampuni ya mawasiliano za Motorola, kuundia kamusi ya Kiswahili-Kifini-Kiswahili iliyotokana na kongoo ya HSC (Helsinki Corpus of Swahili). SALAMA pia imetumika katika utathmini wa kikompyuta wa makamusi kadhaa ya Kiswahili ambapo ripoti zake hupatikana kwa Kiswahili na Kiingereza katika tovuti ya SALAMA na pia katika kujengea upatanisho wa sarufi wa Biblia na vitabu vya Apokrifa kwa Kiswahili kupitia kwa vidahizo vya mizizi ya maneno na uainishaji wa aina za maneno. Kupitia kwa Kituo cha shughuli za Kisayansi kwenye Kompyuta chs Chuo Kikuu cha Kikuu cha Helsinki (Centre for Scientific Computing) katika tovuti www.csc.fi, SALAMA na kongoo yake inapatikana kwa matumizi ya watafiti wanapopeleka maombi ya kuruhusiwa kuitumia. Rekodi na habari kamili kuhusu mradi wa SALAMA inapatikana katika <http://www.njas.helsinki.fi/salama/impl-appl.html> na katika kazi kadhaa za mwasisi wake Arvi Hurskainen ikiwa ni pamoja na <http://www.njas.helsinki.fi/pdf-files/vol13num3/hurskainen.pdf>. Mipango ya baadaye ya mradi huu ni pamoja na kuiboresha tafsiri-mashine ya Kiswahili hadi Kiingereza iliyopo na kuanzisha tafsiri-mashine ya Kiingereza hadi Kiswahili, kutumia mbinu za kikanuni na kitakwimu iwezekanavyo ili kushughulikia utata wa

kisemantiki pale ambapo kanuni za kutegemeka haziandikiki. Pamoja na tafsiri za matini-kwa matini kuna mpango wa kuanzisha programu za kutambua maneno ya lugha kutoka kwa sauti wakishirikiana na wataalamu wa sauti (Hurskainen, 2009).

Haya yote yaliwezekana na yanaendelezwa kutokana na kuwepo kwa hazina ya maneno ambayo ni kongoo/kopasi ya HCS yenye maneno milioni 12.5. Mradi huu wa kuunda kongoo uliofanywa kwa ushirikiano wa kiasi fulani na Chuo Kikuu cha Dar es Salaam kilichokusanya matini nyingi tu kutoka kwa hotuba, mazungumzo na vitabu vya fasihi. Pia kongoo hii imejengwa na magazeti kadhaa ya Kiswahili kutoka Tanzania na machache kutoka Kenya kama vile Al-nur, Majira, Mzalendo, Nipashe, Rai, Uhuru, Kasheshe, Lengo, Alasiri na mengineyo pamoja na tovuti ya habari ya idhaa ya Deutsch Welle ya Kiswahili kutoka Ujerumani kama vyanzo vya kongoo hii iitwayo HCS yaani Helsinki Corpus of Swahili. (Hurskainen, 2004).

1.1 TATIZO LA UTAFITI

Karne ya 21 imenasibishwa na matumizi ya kompyuta katika nyanja mbalimbali za maisha ili kumrahishishia binadamu kazi yake na pia kupata matokeo yenye mashiko zaidi. Watafiti wa lugha ulimwenguni wamehusisha teknolojia hii pia katika kuzichunguza lugha mbalimbali kama tulivyodokezea hapo awali. Lugha za Kiafrika kikiwemo Kiswahili kwa muda mrefu zimewekwa katika kundi la lugha ambazo zina hazina haba ya kikompyuta ulimwenguni (Muhirwe 2007:123) na hivyo kuna haja ya watafiti wa Kiswahili kuzinduka na kufuata mkondo huu. Utafiti huu ungependa kutambua ukweli wa hali ya utafiti wa Kiswahili katika jukwaa la teknolojia ya lugha. Kwa lugha yenye hadhi na matumizi mapana kama Kiswahili, ingepaswa kuwa na miradi mingi tena mikubwa mikubwa ya kiutafiti wa kikompyuta katika taasisi za kiutafiti na vyuo. Hali hii ni tofauti hata kama angalau miradi kama SALAMA inaonyesha kuwa hilo linawezekana. Kwa kutumia mradi wa SALAMA kama kielelezo, utafiti huu ulipenda kuonyesha njia katika uboreshaji wa utafiti wa lugha hii kwa kutumia teeknolojia ya lugha.

Utafiti huu umefaulu katika kuangazia mbinu zilizotumika katika kuchanganua na kufanyia utafiti lugha ya Kiswahili katika Mradi wa SALAMA. Shughuli muhimu imekuwa ni kuonyesha jinsi mbinu hizi zinavyomudu suala zima la kushughulikia data asilia ya lugha ya Kiswahili na hivyo kukuza utafiti wake. Mradi wa SALAMA kama tulivyotangulia kusema unatumia mbinu za Teknolojia ya Lugha yaani programu zilizojengwa katika kompyuta na kuwekewa misingi ya kanuni za kiisimu za kutambua maneno ya Kiswahili kama data bainifu na kuichanganua na kuifafanua kwa mujibu wa maagizo ya kikompyuta (Hurskainen, 2004). Aidha karne ya 21 pia imeshuhudia maendeleo makubwa katika Teknolojia ya Lugha. Teknolojia hii imetumiwa katika Utafiti wa Lugha mbalimbali zikiwemo Lugha za Kiafrika. Kiswahili ni mojawapo wa lugha za

Kiafrika, zilizojikuta katika majukwaa ya kiteknolojia na katika kompyuta kwa miaka mingi sasa. Ni dhahiri kuwa maendeleo ya Kiswahili hayaendi sambamba na maendeleo ya Teknolojia ya Lugha katika ulimwenguni (Hurskainen 2005:125) huu licha ya kuwa lugha hii sasa inatafitiwa katika vyuo vingi na kufundishwa barani Afrika na ng'ambo katika bara la Ulaya, Asia na Marekani. Huu si utafiti wetu kwa sasa. Letu katika kazi hii, na ambalo, tunaamini limeongeza thamani katika taaluma hii, ni uchunguzi wa jinsi wataalamu na watafiti wa Lugha ya Kiswahili wanatumikiza mbinu za kikompyuta katika Teknolojia ya Lugha katika Utafiti wao na ni kwa manufaa gani kwa utafiti wa lugha ya Kiswahili.

Kama tulivyofafanua hapo awali, mradi wa SALAMA ambao ni jukwaa la kiteknolojia ya lugha lenye kuchanganua lugha katika tarakilishi ni ithibati tosha kuwa Kiswahili na lugha nyinginezo za Kiafrika zinaweza kutafitiwa, kuchanganuliwa na kuainishwa kutumia Teknolojia ya Lugha. Kutokana na juhudi hizi, tulitanguliza kujiuliza maswali mengi katika kupata muktadha wa utafiti huu. Ieleweke kuwa haya si maswali ya utafiti huu ila ni ya kutuchokoza fikra. Iwapo mradi huu uliwezekana kwa kufanya yote tuliyotaja, je, ulikuwa na athari gani kwa watafiti wa Kiswahili barani Afrika na kokote ulimwenguni ambapo wapo watafiti wa lugha ya Kiswahili? Mradi wa SALAMA ulishirikisha Chuo Kikuu cha Dar es Salaam kwa kiasi kidogo tu katika uundaji wa Kongoo-matini ya kuchambua Mofolojia ya Kiswahili katika kile kilichoitwa DAHE- Dar-Helsinki Archives (Hurskainen 1999), ila asilimia kubwa iliasisiwa na Prof. Arvi Hurskainen na kuendelezwa Ulaya katika Chuo Kikuu cha Helsinki-Finland. Je, bara la Afrika lina uwezo wa kuvijengea vyuo vyake mazingira ya kiteknolojia ya kufanyia utafiti lugha zao kimitambo, kutumia tarakilishi? Kwa mujibu wa Gikambi (2015) Chuo cha Dar es Salaam hasa Taasisi ya Taaluma za Kiswahili inapaswa kuwa angalau iliendeleza mwanga huu wa kikompyuta ulioanza kwao ila athari hizo hazikubakia au waliosalia nao hawakuumakinikia. Je, programu za kufanyia utafiti na kuchanganulia lugha kiteknolojia kama zile zilizotumika katika mradi wa SALAMA, kwa nini hazivumi kimatumizi barani katika karne hii ya kidijitali?

Hatua kubwa na za kipekee zilizopigwa katika utafiti wa lugha ya Kiswahili kiteknolojia katika mradi wa SALAMA tangu miaka ya 1985, zingekuwa zimenawiri sana angalau katika vyuo vya Afrika Mashariki kulikotokea Kiswahili. Ukweli kuwa juhudi kama hizi hazionekani wala kusikika (Hurskainen 2005) na pia hazijaandikiwa na wataalamu katika vyuo vingine barani vinavyoshughulikia utafiti wa Kiswahili na Lugha za Kiafrika, ni jambo la kufikirisha. Swali ambalo tunaweza kujiuliza tena ni; ni vipi wataalamu na watafiti wa Kiswahili wanaweza kutumia mbinu za kiteknolojia ya kompyuta zilizotumika katika mradi wa SALAMA katika kufanyia utafiti lugha ya Kiswahili badala ya kutegemea tu utafiti uliojikita katika misingi ya king'amuzi (intuition)

au kihisia inayoweza kuitwa mbinu za kimapokeo

Isitoshe, ni dhahiri kuwa machapisho ya kitaalamu ya Isimu-Kompyuta, Isimu-Kongoo na pia Uchakataji wa Lugha Asilia au Teknolojia ya Lugha katika Kiswahili na kuhusu Kiswahili, ni machache sana na pia adimu hata kwenye mitandao. Kwa nini hali iwe hivyo katika jamii ambayo anaendeshwa na teknolojia na hasa tunaporejea lugha yenye wazungumzaji zaidi ya 150 milioni na ambayo ndiyo lugha ya kipekee ya Kiafrika iliyo kati ya lugha rasmi za Muungano wa Afrika?

Suala la kushirikishwa kwa Kiswahili katika matumizi yake kwenye kompyuta na kama lugha ya mawasiliano kwenye vipindi vya media, tovuti na bidhaa za mtandaoni za kampuni kubwa kubwa kama Google, Microsoft, Linux, Facebook, Firefox, Opera Mini na kadhalika na kwenye mashirika ya habari kama BBC, Deutsch Welle-Swahili, VOA na sasa katika Swahilihub.com na kadhalika, ni ithibati tosha kuwa hii ni lugha iliyo na uwezo wa kujitana kiteknolojia na kikompyuta katika nyanja zote (Mabeya 2009).. Je, watafiti na wataalamu wa Kiswahili wanaendelea kushiriki kwa kiasi gani katika michakato ya ujenzi wa mitandao hii ya Kiswahili na wenyewe wana mikakati gani ya kushirikisha lugha hii katika majukwaa zaidi ya kikompyuta ili kusaidia utafiti wake?

Hoja kuu inayosukuma mawazo haya na mkururo huu wa maswali ni kuwa, mradi wa SALAMA umedhihirisha kuwa teknolojia ya lugha katika kompyuta inarahisisha utafiti na uchanganuzi wa Lugha ya Kiswahili, Hurskainen (2008). Lile ambalo si dhahiri ni jinsi wazo na mbinu hizi zilivyosambaa na kwa kiasi gani katika idara za Kiswahili katika vyuo na taasisi barani Afrika na hata Ulaya, Asia na Marekani kunakotafitiwa Kiswahili. Utafiti huu umeoyesha natija za mkabala huu mpya wa utafiti na kupendekeza haja ya kuufumbata kama jumuiya ya watafiti wa lugha ya Kiswahili kwa kufafanua kwa uwazi mifano na miktadha ya matumizi ya mbinu hizi zilizotumiwa na SALAMA na kuangazia ufanisi wake na kupendekeza jinsi ya kupiga hatua katika taaluma hii.

Isitoshe, suala la Isimu-Kompyuta au Teknolojia ya Lugha huhitaji kongoo yenye maneno mengi au vipashio vingi vya leksika iwezekanavyo (Hurskainen 2008) ili iwe inawakilisha hali halisi ya lugha.

Kwa kuangalia jinsi mradi wa SALAMA ulivyotekeleza michakato mingi ya data za Kiswahili kutumia kongoo ya maneno 12.5 milioni pekee, ni dhahiri kuwa kongoo kubwa na shirikishi zaidi inaweza kujengwa kwa kuwa sasa Kiswahili kina machapisho mengi hata ya mtandaoni ambayo yanaweza kutengeneza kongoo ya maneno takribani nusu bilioni (500,000,000) au hata zaidi kwa urahisi sana. Magazeti ya Kiswahili nchini Kenya na Tanzania na vitabu vya Kiswahili vilivyowahi kuchapishwa hapa Afrika Mashariki pekee kwa pamoja vinaweza kuwa na kongoo ya zaidi ya maneno bilioni moja (Gikambi, 2015). Ni sadfa kubwa kuwa ²Taasisi ya Lugha za Kiafrika (ACALAN) ambayo ni kitengo cha Lugha cha AU- Muungano wa Afrika kilianza juhudi za

kushirikisha Kiswahili katika tafiti za kiteknolojia na katika Isimu Kompyuta kwa kuandaa warsha ya kufundisha wataalamu wa Kiswahili jijini Dar es Salaam mnamo Machi 2015 ili kuanza mchakato wa kuunda kongoo kubwa zaidi ya Kiswahili duniani wakiwa na lengo la kongoo ya maneno 100 milioni. Kwa mujibu wa Prof. Kimani Njogu na Dkt Suleiman Sewangi wanaoongoza mradi huu, tayari taasisi hii imefanikiwa kukusanya Kongoo-matini yenye maneno zaidi ya nusu milioni. Katibu Mkuu mtendaji wa Taasisi ya Lugha za Kiafrika –ACALAN, Prof Matsinhe Francisco Sonzinho (2015) (katika ²pendekezo lake kwa kampuni ya Nationa Media Group na chama cha Wachapishaji nchini Kenya (KPA) ameomba ushirikiano wao kwa kuazima matini za Kiswahili za kuundia kongoo hiyo kutoka kwa magazeti, tovuti na vitabu kwa usanjari huo. Muungano huu unatumia Kiswahili kama lugha yake ya pekee rasmi ya Kiafrika na kukiendeleza ili kiwe lugha ya ³kuunganisha watu nje ya mipaka ya mataifa binafsi ya Afrika Mashariki na kati. Miradi kama hii ya kiteknolojia ya lugha inahitaji kutafitiwa ili iwekwe katika muktadha wake unaoenda na wakati na hilo pia linahalalisha utafiti huu.

Mwisho, utafiti huu utanua kujaza pengo la kiutafiti kuhusu matumizi ya programu za kuchakata data ya lugha ya Kiswahili katika kompyuta. Tafiti zilizotangulia na kuandikwa kwa Kiswahili zimeonyesha jinsi kompyuta kama chombo imetumia lugha ya Kiswahili na kuisaidia kuenea kama wanavyojadili Kamau, 2005 na Mabeya, 2009. Kazi hii imeangalia jinsi mbinu zenyewe zinavyokuza utafiti wa lugha ya Kiswahili wala si uwepo wa Kiswahili kwenye mashine yenyewe au mtandaoni. Ni wazi kuwa teknolojia na teknolojia ya lugha ni mambo mawili tofauti na hivyo utafiti huu utakuwa na mchango mkubwa katika taaluma inayoinukia katika utafiti wa lugha.

1.2 MALENGO YA UTAFITI

Utafiti huu una malengo yafuatayo:

- i) Kubainisha mbinu na programu za kikompyuta zilizotumika katika Utafiti na Uainishaji wa Lugha ya Kiswahili katika Mradi wa SALAMA.
- ii) Kuchunguza jinsi mbinu zilivyotumika katika Utafiti wa Kiswahili kwa kutumia Teknolojia ya Lugha kwenye mazingira ya Mradi wa SALAMA.

.....
²Taasisi ya Lugha za Kiafrika (ACALAN)- African Languages Academy ina makao makuu yako BAMAKO, Mali na imeshikiri katika kuunda mikakati ya kuinua Kiswahili barani kama lugha rasmi.

³pendekezo... Taasisi ya ACALAN kupitia Katibu Mkuu wake Prof Matsinhe Francisco Sonzinho anayeondoka alimshirikisha mtafiti wa kazi hii katika uundaji kongoo hii. Barua ya kuomba ubia kati ya Nation Media Group na Chama cha Wachapishaji (KPA) Kenya Publishers Association zinashughulikiwa kwa mujibu wa katibu huyo.

iii)Kutathmini mchango wa mbinu na programu hizi za Mradi wa SALAMA katika Utafiti wa Lugha ya Kiswahili kwa kutumia Teknolojia ya Lugha katika siku za usoni na zinavyoweza kufaidi taaluma ya Kiswahili.

iv) Kutathmini mchango wa Kongoo katika utafiti wa kiteknolojia ya lugha ya Kiswahili katika Mradi wa SALAMA sasa na katika siku za usoni.

1.3 MASWALI YA UTAFITI

Utafiti huu ulikuwa unalenga kujibu na umejibu maswali yafuatayo;

- i) Ni mbinu gani zinazotumika katika kufanyia utafiti lugha ya Kiswahili kutumia teknolojia ya lugha kikompyuta kwenye Mradi wa SALAMA?
- ii) Ni vipi mbinu hizo zilitumika katika Utafiti wa Kiswahili kwa kutumia Teknolojia ya Lugha kwenye mazingira ya kikompyuta ya Mradi wa SALAMA?
- iii) Ni mchango gani unaoletwa na mbinu au programu hizi za SALAMA katika Utafiti wa Lugha ya Kiswahili kwa kutumia Teknolojia ya Lugha katika siku za usoni na zinaweza kufaidi vipi taaluma ya Kiswahili?

iv) Kongoo ya Kiswahili inasaidiaje katika utafiti wa kiteknolojia ya lugha katika Mradi wa SALAMA sasa na katika siku za usoni?

1.4 SABABU ZA KUCHAGUA MADA HII

Tulipata motisha wa kuchagua mada hii kwa sababu kadhaa. Mwanzo, mada ya teknolojia ya lugha ni mada ya kisasa na inaendana na wakati kama tulivyotanguliza kusema. Teknolojia ya kompyuta inatumika katika kila nyanja ya maisha ya binadamu katika karne hii ya 21 na hivyo hatuna budi ila kuishirikisha katika utafiti wa lugha ya Kiswahili. Tayari lugha ya Kiswahili iko katika matumizi ya kiteknolojia tangu ilipoanza kutumiwa katika mitambo ya benki, kampuni za simu na kampuni za huduma za kompyuta kama Microsoft East Africa, Linux, Google na kadhalika (**Mabeya, 2009**). Ushahidi wa usasa huu unaonekana katika juhudi za wadau wakuu wa Kiswahili barani Afrika na ng'ambo. Kufikia mwaka wa 2011, Taasisi ya Lugha za Kiafrika –ACALAN yenye makao yake mjini Bamako nchini Mali, chini ya Muungano wa Afrika (AU) ilikifanya rasmi Kiswahili kuwa lugha ya mazungumzo nje ya mipaka ya kitaifa na kukiundia kamisheni ⁴KVCBLC) yake iliyo na wadau kutoka kila taifa katika Jumuiya ya Afrika Mashariki.

.....
⁴KVCBLC – (Kiswahili Vehicular Cross Border Language Commission) yaani Kamisheni ya Kiswahili kama lugha ya kutumika nje ya mipaka ya mataifa ya kanda ya Afrika Mashariki iliundwa jijini Nairobi (2011) na inafadhiliwa na Taasisi ya ACALAN ili ikuze Kiswahili katika kanda ya Afrika Mashariki. Kulikuwa pia na lugha nyingine kuu tano barani Afrika.

Hili linatokea katika kanda ya Afrika Mashariki na Kati, miaka kadhaa baada ya muungano huo kukikweza Kiswahili kuwa lugha rasmi ya Kiafrika ya pekee katika muungano wa AU.

Msukumo wa kutumia teknolojia ya lugha katika karne hii umepelekea taasisi hii kuanzisha rasmi mchakato wa kuunda kongoo kubwa zaidi ya lugha ya Kiswahili kwa matumizi ya utafiti wa lugha hii kikompyuta kwa kuandaa warsha ya kuwafundisha wanaisimu kongoo na wataalamu wa Kiswahili katika masuala ya uundaji wa kongoo mnamo mwezi wa Machi 2015 kama tulivyotaja.

⁶(Matsinhe, 2015). Juhudi hizi zote zinaonyesha mwamko mpya katika mtazamo huu wa kikongoo na ushirikishaji wa teknolojia ya lugha katika utafiti wa lugha ya Kiswahili. Utafiti umepata mashiko katika mwamko huu mpya kwa kuwa matumizi ya teknolojia ya lugha yanategemeka zaidi kuliko matumizi ya ung'amuzi wa kimapokeo katika kufanyia Kiswahili utafiti.

Pili, mbinu za kisayansi na kijarabati zimetumika na wataalamu waliotangulia kufanya utafiti katika eneo hili kwa hivyo utafiti wa mada hii haubabaishi na kubahatisha bali unakwenda na data za kuonekana na kuaminika. Watangulizi wangu katika eneo hili la teknolojia ya lugha wanaona kuwa matumizi ya kompyuta katika kuchambua lugha za Kiafrika yataharakisha mchakato wa uchambuzi huo (Hurskainen, 2008:42) na pia yatafaidi utafiti katika uundaji wa mbinu za kikompyuta kama vile mbinu za kusahihisha lugha kwa mfano vikagua tahajia, n.k., mbinu za utungaji wa kamusi na nyingine nyingi ambazo zote zinaweza kutumika kutafuta data mbalimbali katika kongoo kwa ufasaha mkubwa, wazo linaloungwa mkono na Sewangi 2000, 2001, Hurskainen 2003 na Muhirwe 2007, kuliko mbinu za kubahatisha dhana kwa kutumia data haba za lugha.

Uchanganuzi wa kongoo-matini yenye idadi ya mamilioni ya maneno kutumia programu ya kompyuta kwa dakika chache, inaibua data za kuaminika zaidi kuliko kuchambua sampuli ndogo ya data ya wazungumzaji yenye maneno machache. Utafiti huu unafaa ili kuonyesha njia katika hatua hizi mpya zinazoonyesha kuwa na natija na kasi zaidi na hivyo za kutegemewa zaidi. Hata hivyo, ni muhimu tufahamu kuwa kuna nafasi ya binadamu katika michakato yote hii na hivyo hatutarajii kuwa kompyuta hufanya kila kitu yenyewe (Hurskainen 2009).

Tatu, kuna tafiti zilizofanywa na watangulizi wangu kama Kamau (2005) aliyetafitia matumizi ya Kiswahili katika tarakilishi na akataja jinsi mradi wa SALAMA ulivyotumia Kiswahili katika programu zake. Alipendekeza utafiti zaidi wa kina ufanywe kuhusu jinsi Kiswahili kinavyoweza kutumiwa katika tarakilishi. Pia mwenzake Mabeya (2009) aliyefanya utafiti wake kuhusu uundaji wa istilahi katika mradi wa Microsoft-East Africa amependekeza kuwa lugha ya Kiswahili imefikia kiwango cha kwenda sambamba na teknolojia ibuka ya tarakilishi na hivyo inafaa kufanyiwa utafiti

zaidi. Amesema kuwa kuna haja ya kuangalia miradi inayoweza kusaidia katika utafiti wa Kiswahili kiteknolojia kwa kuwa alionelea mradi wa Microsoft East Africa kama mkakati tu wa biashara ya kimataifa ya kuuza bidhaa ororo au programu zake na wala si juhudi zilizolenga kukuza utafiti wa Kiswahili. Bila shaka mkabala wetu huu unakuza utafiti wa lugha ya Kiswahili kwa kuwa unamakinikia lugha haswa.

Isitoshe, utafiti huu ulichagua kuangazia mradi wa SALAMA kwa sababu huu ndio mradi wa kipekee ulimwenguni mzima ambao unashuheni uanuwai wa programu za teknolojia ya lugha zinazoshughulikia lugha ya Kiswahili kikamilifu. Katika SALAMA, kama tulivyotanguliza, kuna programu za kuchanganua Kiswahili kimofolojia, kisintaksia, kisémantiki, kuna mbinu za uundaji kamusi, kuna programu za kutafsiri, kuna vikagua tahajia, programu za kutatua utata wa kimaana katika lugha, kufundisha Kiswahili na hata kuwepo kwa kongoo-matini ya maneno zaidi ya milioni 12.5 ambayo inaweza kuwa mali ghafi ya utafiti wa lugha ka jumla. SALAMA, kwa hivyo, ni mazingira mwafaka zaidi ya kuchunguza jinsi teknolojia hii inavyoweza kuongeza thamani katika utafiti wa Kiswahili. Kwa kuchunguza na kuathmini mradi wa SALAMA kama mazingira mwafaka ya kukishughulikia Kiswahili, tuliibuka na mifano bora ya ufanisi wa teknolojia ya lugha hii na mikakati ya kuangazia katika mustakabali wa mambo.

Mwisho, kushirikishwa kwa mtafiti wa kazi hii katika mradi wa Taasisi ya Lugha za Kiafrika-ACALAN katika uundaji wa kongoo ya Kiswahili mapema mwaka huu kama mmoja wa miradi ya Muungano wa Afrika-AU ya kukuza lugha za bara hili, kulikuwa ni mshawasha tosha wa kuendelea kufanyia utafiti masuala yanayohusiana na Kongoo na teknolojia ya lugha. Utafiti huu ulifaidika pakubwa na juhudi hizo ambazo tayari zimezianza.

Kutokana na sababu hizi, ni wazi kuwa utafiti huu unatazamiwa kuchangia katika tafiti zilizofanywa hapo awali kuhusu matumizi ya Kiswahili katika kompyuta. Utafiti huu unatazamia kujaza pengo lililoachwa na watafiti waliotangulia kwa kuwa hatujakutana na tafiti zilizoandikwa kwa Kiswahili zinazomakinikia kwa kina na kwa ukamilifu wake vielelezo vya TL kama mradi wa SALAMA kwa jumla au matumizi ya teknolojia ya lugha na nafasi ya mradi huu katika utafiti wa lugha ya Kiswahili hasa.

Isitoshe utafiti huu umejaza pengo la kuonyesha jinsi nadharia kama Msambao wa Ugunduzi na nadharia ya Isimu Kongoo zinavyoweza kutumiwa katika uchunguzi wa teknolojia ya lugha ya Kiswahili ambayo ni lugha ya Kiafrika. Nadharia hizi mbili zimetiana nguvu kwa kuwa ile ya Msambao wa Ugunduzi inatalii kipengele cha teknolojia ya kikompyuta jinsi ujuzi huo unavyorahisisha mambo katika michakato ya kushughulikia lugha kiteknolojia. Mtazamo wa

kikongoo unafidia kwa kuleta upande wa uhalisia ya lugha katika matumizi yake. Kongoo ni mkusanyiko wa matini halisi ya lugha kama ilivyotumiwa katika mazungumzo au maandishi (Hurskainen 2005). Nadharia ya Msambao wa Ugunduzi imewahi kutumiwa kufanyia utafiti masuala ya Kiswahili na matumizi yake katika kompyuta ila hatujakutana na mahali ambapo zote mbili zimewahi kutumika tena katika utathmini wa matumizi ya teknolojia ya lugha yaani programu zinazochanganua lugha kijarabati kwa pamoja. Bila shaka basi, utafiti huu umekuwa na upekee wake katika kutoa mchango mkubwa kwa taaluma ya Kiswahili hasa katika enzi inayoelekea kukumbatia teknolojia ya tarakilishi kwa utafiti wa kila aina. Isitoshe, nadharia ya Isimu Kongoo imeendelea kupata umaarufu zaidi kama mtazamo wa utafiti wa lugha unaojumuisha vipengele vyote kutoka fasihi, isimu, diskosi, isimu jamii na kadhalika kwa kuwa ni kusanyo la lugha asilia.

1.5 UPEO NA MIPAKA YA UTAFITI

Utafiti huu ulijifunga katika tathmini ya programu za kikompyuta zilizotumika katika mradi wa SALAMA kushughulikia Kiswahili na jinsi zinavyochangia katika utafiti wa lugha hiyo.

Kiswahili kimeundiwa kongoo/kopasi mbalimbali ambazo zimetumiwa katika utafiti, uainishaji na uchanganuzi wake kikompyuta kutumia stadi inayoitwa Teknolojia ya Lugha au Mchakato wa Lugha Asilia. Utafiti huu, kwa hivyo umechunguza teknolojia ya lugha katika Mradi wa SALAMA ambao ni mazingira ya kikompyuta ya kushughulikia lugha ya Kiswahili ulioasisiwa katika Chuo Kikuu cha Helsinki Finland. Mtafiti alimulika kwa kina baadhi ya jitihada zilizofanywa na mwasisi wa mradi huu Prof. Arvi Hurskainen na wataalamu wenzake katika kuunda mazingira ya SALAMA kwa kuvumbua programu mbalimbali ambazo zinatumiwa katika kushughulikia Kiswahili kiutafiti na katika uchanganuzi wa vipengele vya kiisimu na lugha kwa jumla.

Katika kufanya hivi, uchunguzi huu umeegemea katika utathmini wa mbinu zilizotumika katika Utafiti na Uainishaji wa Lugha ya Kiswahili katika Mradi wa SALAMA na mchango wake katika kukuza stadi ya utafiti wa lugha hii. Programu za kuchanganua lugha kimofolojia, kisintaksia na kisemantiki zilizosanidiwa na SALAMA pamoja na matumizi yake katika uundaji kamusi, utumiwaji katika tafsiri-mashine (tafsiri-oto), vikagua-tahajia, ufundishaji wa lugha na vitatuzi vya utata wa maana ya lugha zimeangaziwa na kutathminiwa. Aidha utafiti huu umeangalia jukumu la kongoo ya SALAMA-HSC (Helsinki Corpus of Swahili) na matumizi yake katika utafiti wa kiteknolojia na jinsi inavyokuza na kuhifadhi lugha ya Kiswahili.

Ni muhimu kutaja kuwa, uchunguzi huu, umeangazia Mradi wa SALAMA kama kielelezo au kifani cha mazingira ya kikompyuta na jinsi Teknolojia ya Lugha imetumiwa kufanyia utafiti na kuchanganulia Kiswahili kiisimu. Kwa hivyo, uchunguzi huu haukujihusisha na uchanganuzi wa

data ya Kiswahili kutumia programu zozote za kompyuta zilizotumiwa na Prof. Arvi Hurskainen katika SALAMA au zinginezo zozote zile. Hii ni kwa sababu hilo halikuwa lengo la utafiti huu na letu lilikuwa kutafuta ithibati ya jinsi zilivyofanya kazi na manufaa yake katika utafiti wa Kiswahili. Kwa hivyo kulikuwa hakuna haja ya gharama ghali za kusafiria hadi Helsinki, Finland kuweza kutangamana na baadhi ya programu hizi au kung'ang'ana kuruhusiwa kuzitumia mtandaoni. Mchakato wa kuruhusiwa kuzitumia mtandaoni pia una uchangamano ambao hauhitajiki kwa kiwango hiki cha masomo. Badala yake, uchunguzi huu uliangalia tu jinsi mbinu na programu mbalimbali za kuchakata data ya lugha halisi kikompyuta na kuiainisha zilivyotekelezwa katika mradi huu na jinsi zinavyoweza kufaidi utafiti wa Kiswahili. Utafiti huu haukuchanganua kongoo ya HCS kutumia mbinu na programu za SALAMA ila uliangalia mifano ya jinsi programu zenyewe zilivyoshughulikia lugha ya Kiswahili katika mradi wa SALAMA kwa mujibu wa ripoti 19 za Hurskainen zilizo katika tovuti ya SALAMA na makala nyingine za kitaalamu kama zitakavyobainishwa katika kipengele cha ukusanyaji wa data.

Mwisho, kwa kurejelea misingi ya kinadharia katika utafiti huu, mtafiti aliegemea kwenye vipengele maalum vya nadharia zilizotajwa katika utafiti huu yaani (Nadharia ya Msambao Ugunduzi na Nadharia ya Isimu Kongoo). Licha ya kuwa Hurskainen (2008:42-44) anaonyesha jinsi nadharia za kiisimu na kihisabati zinavyoweza kutumiwa katika ujenzi wa mbinu za kikompyuta na pia jinsi teknolojia hii ya lugha inavyosaidia katika upimaji na utathmini wa ukamilifu wa nadharia husika, utafiti huu haukujihusisha na nadharia hizo za kihisabati kwa kuwa masuala tunayomakinikia, licha ya kuwa yanaangalia lugha katika mazingira ya kialgorithmu, utafiti wenyewe haupimi masuala ya kanuni za kiisimu au kihisabati. Kazi hii haikujihusisha na masuala ya vipengele mahususi vya kiisimu vilivyotajwa na hivyo hatukuvizamia. Badala yake utafiti huu umejibana tu katika kuangalia jinsi mbinu hizi za kikompyuta zinazochanganua lugha kiisimu au kimuundo katika mradi wa SALAMA zinavyotoa mchango katika uwanja wa kufanyia utafiti lugha ya Kiswahili. Maoni ya mwasisi wa mradi wa SALAMA Arvi Hurskainen pamoja na yale ya wanafunzi wake waliushughulikia mradi huu, yalikuwa muhimu sana kwa utafiti huu na hivyo tuliweza kubaini hali halisi ya mradi huu sasa hivi, miaka 30 baada ya kuasisiwa. Kwa sababu za gharama ya usafiri na matumizi mengine, mtafiti hataweza kusafiri hadi Helsinki Finland kwa mahojiano au kujaribisha mbinu hizo kwenye kompyuta zao, ila alitegemea majibu ya hojaji kwa barua pepe na mahojiano kwa simu, pamoja na kusoma, vitabu, ripoti na makala yanayoeleza hali ya mradi wa SALAMA katika tovuti yake na maktabani.

1.6 UDURUSU WA YALIYOANDIKWA KUHUSU MADA HII

Utafiti unaohusu teknolojia ya lugha katika uchambuzi na utafiti wa lugha hasa lugha za Kiafrika kama Kiswahili ni dhana ya kisasa mno. Dhana hii haijapata kuandikiwa na watu wengi kwa kuwa ndipo inachipuza na kujidhihirisha. Wapo watafiti wengi waliomakinikia lugha za kimataifa kama Kiingereza, Kifaransa, Kijerumani na kadhalika katika kompyuta kwa kuwa matini zake zilipatikana katika muundo wa kusomeka kwa kompyuta kwa muda mrefu hadi sasa (De Pauw and de Schryver 2008). Wataalamu hawa walionyesha jinsi lugha ya Kiswahili ilivyo kifua mbele katika baadhi ya lugha za Kiafrika kwa kuwa hata inapatikana mtandaoni na sasa iko katika nafasi nzuri ya kuchunguzwa kutumia teknolojia ya lugha, mawazo yanayoungwa mkono na Muhirwe (2007:124) katika utafiti wake unapendekeza mikakati ya kushirikisha lugha za Kiafrika kikiwemo Kiswahili katika utafiti wa kutumia zana na majukwaa ya kikompyuta.

Hurskainen (1992), aliyeasisi mradi wa SALAMA aliona kuwa kuna ongezeko la hamu ya matumizi ya kompyuta katika utafiti na ufundishaji wa lugha za Kiafrika. Iwapo kazi yake ilionyesha kuwa lugha za Kiafrika kama Kiswahili ziliafiki kwa kila njia kutafitiwa kwa michakato ya kikompyuta, alisema kuwa, miaka ya 1980 na 1990 juhudi nyingi zinazohusu Isimu Kompyuta zilisisitizwa na kuelekezwa katika miradi ya kuunda vitafsiri-oto (programu za tafsiri-mashine) na pia programu za kuchambua lugha kisintaksia na kusema kuwa lugha za Kiafrika ziliwapa watafiti changamoto za kutosha katika uainishaji wa aina za maneno na akatoa mfano wa jinsi lugha za kibantu zilifaa kujaribisha programu za kuchambua lugha kimofolojia kwa kuwa zina viambishi na minyambuliko mingi.

Hurskainen (1999:145), akifafanua mazingira ya SALAMA, alieleza jinsi hazina ya lugha ya Kiswahili katika kompyuta ilivyosaidia katika upatikanaji wa data ya lugha kwa urahisi na ufasaha hasa pale maneno yalikuwa yameainishwa kimofolojia, kisantiki na kisintaksia na hivyo ukawa ni mfumo uliofaa sana katika uundaji wa programu za utafiti na matumizi ya kawaida kama kuunda kamusi na kukagua tahajia ya lugha ya Kiswahili.

Akiyanukuu maoni ya Sewangi (2002), Mabeya (2009) anasema kwamba, tarakilishi ina matumizi mapana sana na lengo la msingi ni kuhifadhi data na kurahisisha kuichambua. Anafanua kuwa kongoo matini (text corpus) hutumika kama hifadhi ya lugha katika tarakilishi ili iweze kuchanganuliwa kwa malengo mbalimbali ya kiutafiti. Katika Kiswahili kuna programu ya kukagua usahihi wa maandishi-kikagua tahajia (spell checker) na kongoo matini ya Kiswahili yenye zaidi ya maneno milioni kumi na mbili ya Helsinki Corpus of Swahili (HCS) katika mradi wa SALAMA. Utafiti wetu pia utaangazia teknolojia ya lugha na jinsi kongoo ya HCS ilivyotumika katika mazingira hayo ya kikompyuta. (Sewangi, 2001) ila tofauti na mwenzetu huyu hakuangalia uundaji

wa utoaji wa misamiati ya taaluma na sajili mahsusi ila tunaangalia teknolojia ya Kiswahili kwa jumla katika utafiti wa Kiswahili. .

Mohochi (2002) anazungumzia mabadiliko katika ulimwengu kutokana na uvumbuzi wa TEKNOHAMA na kuongeza kuwa utafiti wa Kiswahili unaweza kufaidika pakubwa na mtandao kwa kuhifadhi matokeo ya utafiti uliofanywa, ili matokeo haya yawafikie wale ambao hawafahamu lugha ya Kiingereza. Lugha ya Kiswahili haina budi kufanywa lugha ya matumizi katika tarakilishi kwa kuzalisha msamiati wa tarakilishi wa Kiswahili. Utafiti huu uliangalia kwa kiasi fulani jinsi mtandao unavyotumika kujenga kongoo kama mali ghafi ya Teknolojia ya Lugha na si mtandao kama uwanja wa uzalishaji wa misamiati

Kihore (2002) anaeleza kwa jumla jinsi lugha ya Kiswahili inavyoweza kuingizwa na kutumiwa katika mfumo wa teknolojia ya habari, kama ilivyofanywa katika lugha nyingine kuu duniani. Anaendelea kujadili kuhusu kuingizwa kwa lugha hii katika mfumo wa kompyuta ambao unamruhusu mtumiaji wa lugha hii kuona na kufanya mambo yote ya kompyuta katika Kiswahili. Pia kuandaa programu au kuwezesha zile zilizopo kuchanganua data za lugha hii ili ziweze kuwa na matokeo yanayoweza kutumika katika kazi mbalimbali. Wazo hili la kuchanganua data za lugha tumelikweza kwa kufanyia utafiti programu zinazofanya hivyo katika mradi wa SALAMA kama kifani chetu.

Lacey (2004) alijadili uwezekano wa kutumia lugha moja ya Kiafrika katika kuzihifadhi lugha nyingine na kuziondolea uwezekano wa kutoweka katika matumizi ya tarakilishi. Kwa kuzitumia lugha za Kiafrika katika tarakilishi lugha hizi zitahifadhiwa mbali na kuwa teknolojia hii itawafikia Wafrika wengi iwezekanavyo. Alifafanua kuwa kwa wakati ule kampuni za kompyuta kama Microsoft East Africa tayari zilikuwa zimeanzisha juhudi za kutumia Kiswahili katika tarakilishi ili iweze kuwafaidi wazungumzaji wa lugha mbali na kujiimarisha kibiashara. Hata hivyo, hakuweza kuonyesha jinsi teknolojia ya lugha inavyoweza kusaidia katika juhudi za kuhifadhi lugha katika kongoo, suala tulilolishughulikia katika kazi hii.

Iraki (2004) pia ameligusia suala la Microsoft East Africa kukifanya Kiswahili kuwa mojawapo ya lugha katika programu zake. Anaeleza kuwa kampuni hii imetambua kuwa kuna soko kubwa la huduma zake katika maeneo yanayozungumzwa Kiswahili. Ingawaje hatukufafanua jinsi kampuni inavyotekeleza mradi huu, utafiti huu umefafanua jinsi programu za teknolojia ya lugha zinavyofanya kazi kutumia lugha ya Kiswahili katika mkusanyiko wa kongoo-matini.

Sewangi (2002) ametoa maoni yake kuhusu faida ya tarakilishi kama nilivyozungumzia hapo awali. Kwake yeye lengo la msingi la kutumia tarakilishi ni kuhifadhi data na kurahisisha kuichambua na

nafasi ya kongoo katika jukumu hilo. Utafiti wetu umeimarisha hoja zake zaidi na kuonyesha jinsi teknolojia ya lugha inavyochangia katika eneo la utafiti wa lugha ya Kiswahili.

Ng'ang'a (2005) ameshughulikia tatizo la utatuzi wa utata wa maana ya neno katika muktadha wa tafsiri-mashine za Kiswahili-Kiingereza na hali za kufundishia lugha hii ya Kiswahili. Ameangalia matumizi ya programu za kompyuta katika mifumo ya mtandao wa Bayesian na kushirikisha vielelezo vya mradi wa SALAMA hasa kipengele cha SWATOL katika tasnifu yake ya Uzamifu iliyotalii SALAMA kwa mapana. Kwa kuangalia ngeli za maneno ya Kiswahili kwa muktadha ya kikongoo alichunguza jinsi utata wa kisemantiki wa neno unavyoweza kutatuliwa. Utafiti wetu ulipata thamani kubwa kutoka na utafiti huu kwa misingi ya jinsi mtafiti alivyotumia programu za kufundisha Kiswahili na za kutatua utata ila sisi tumemakinikia mchango wa juhudi hizo katika utafiti wa Kiswahili. Utafiti wake pia uliwasilishwa kwa Kiingereza tofauti na huu wetu unaotumia lugha inayotafitiwa yaani Kiswahili katika kuelezea mawazo hayo.

Kamau (2005) ametafiti juu ya matumizi ya Kiswahili katika tarakilishi. Ameshughulikia miradi kama vile SALAMA, KIKO, Microsoft East Africa, mradi wa kugeuza matini za Kiswahili hadi usemi na mradi wa injini pekuzi ya Kiswahili katika mtandao. Utafiti huu ulichukua mkondo tofauti na mkabala wa kuonyesha tu uwepo wa Kiswahili katika majukwaa haya ya kidijitali na badala yake kuzamia zaidi katika jinsi mbinu mbalimbali za kikompyuta katika mradi wa SALAMA zinavyochangia katika utafiti wa Kiswahili. Kwa upande mwingine, utafiti huu utakuwa unaendeleza kutoka alipoachia Kamau (2005) kwa kuwa alitaja mbinu kadhaa zilizotumika katika mradi wa SALAMA ambao ndio kifani chetu cha utafiti. Kwa hivyo kazi hii ilionekana kama mojawapo wa misingi ya utafiti huu.

Kamau (2005) pia ameangalia jinsi Kiswahili kilivyoshirikishwa katika mradi wa ⁵LLSTI- *Local Language Speech Speech Technology Initiative* wa Chuo Kikuu cha Nairobi katika kompyuta ulioweza kutambua na kugeuza matini hizo kuwa sauti hapa nchini Kenya. Mradi huu uliowashughulisha Wakenya mhandisi Muchemi Gakuru na wataalamu wa lugha; Kamanda Ngugi na mwenzake wa Kifaransa F. N. Iraki ulifaulu kwa kutimiza hilo kwa kukusanya maneno 10,558 kutoka kwa riwaya za Kiswahili, Koran na Biblia. Matini hizo zilisomwa na aliyekuwa mtangazaji wa Nation FM wakati huo, Ken Walibora na sauti ya kila neno kuhifadhiwa kwa kompyuta.

.....
⁵LLSTI- *Local Language Speech Speech Technology Initiative* nradi wa Chuo Kikuu cha Nairobi unaotolewa habari katika <http://www.outsideecho.com/llsti/downloads-languages-kiswahili.html>. Ulihusika na kubadilisha sauti na kuwa matini. (T.T.S yaani Text To Speech).

Kamau anaeleza manufaa ya programu kama hizi kuwa zinaweza kutumika kwa viotomotela vya benki (A.T.M), kuwafundisha wasioona Kiswahili kwa kusikia sauti, kwenye simu na pia kufundisha wageni Kiswahili. Sehemu hii ya utafiti wa Kamau imekuwa msingi bora kwa utafiti huu kwa kuwa tumekuza mawazo haya kwa kuonyesha jinsi teknolojia ya lugha ilivyopiga hatua tangu hapo kwa kuwa mradi wa SALAMA hata haukuhitaji sauti ya mtu kusoma matini na kuhifadhi katika kumbukizi za tarakilishi. Utafiti wetu umeonyesha jinsi teknolojia ya lugha inavyoweza kubadilisha matini kuwa sauti na sauti kuwa matini kwa kutumia programu za kikompyuta za SALAMA ambayo hutambua matini kama vipashio vya kikompyuta.

Utafiti wa aina hii pia uliofanywa na Ng'ang'a (2010) aliposhiriki katika mradi mkubwa kabisa wa kuweka lugha ya Ibo kutoka Nigeria katika sauti na hatimaye katika matini na kuhifadhi kama kongoo-matini. Tushajifunza kuhusu uendeleu wa miradi kama hii kwa kuwa licha ya kuwa mradi wa LLSTI ulikuwa mkubwa na wenye nia njema ya kutumia teknolojia ya lugha katika tarakilishi kufanya raia aliye mashinani hata kama hakusoma kuweza kutumia Kiswahili, ulikwama kutokana na changamoto mbalimbali ila huu wa SALAMA unaendelea vizuri.

Kahigi (2007) naye ameelezea juu ya ujanibishaji wa *Microsoft Office (2003)* na *Windows XP*. Ameelezea misingi ya uundaji wa istilahi ambayo waundaji walikubali kutumia istilahi zilizoundwa. Utafiti wetu haukulengi mbinu za uundaji wa istilahi lakini ulitathamini jinsi istilahi mbalimbali za lugha ya Kiswahili zinakuwa hazina kubwa ya teknolojia ya lugha hasa pale istilahi hizo zinapoingizwa katika kongoo-matini kama lugha halisi.

Mabeya (2009) amechunguza matumizi ya Kiswahili katika mradi wa *Microsoft East Africa* na kubainisha mbinu zilizohusika katika uundaji wa msamiati katika mradi huu. Aidha, umeonyesha jinsi jinsi mbinu hizo zilizotumika zimechangia katika kukikuza na kukiendeleza Kiswahili kama lugha ya teknolojia. Anaona kampuni mbalimbali na miradi kama Microsoft East Africa kama kampuni za kibiashara zinazotumia lugha ya Kiafrika kwa kujieneza kiuchumi ila si kwa lengo la kuzikuza lugha hizo kiutafiti. Utafiti huu ulitujengea hoja ya kuonyesha jinsi mradi wa SALAMA, tofauti na Microsoft, unavyojenga utafiti wa Kiswahili kiteknolojia kwa kukijengea mazingira ya kikompyuta ya kuishughulikia lugha bila miegemeo yoyote ya kibiashara na kiuchumi.

Hurskainen (1992, 1999, 2004a, 2004b 2008, 2009) ambaye ni muasisi wa SALAMA, ameelezea bayana jinsi jukwaa aliloanzisha la teknolojia ya lugha linavyotumia programu kadhaa za kikompyuta kuainisha lugha ya Kifini, Kiingereza, Interlingua na Kiswahili. Katika tafiti zake nyingi zilizoandikwa kwa Kiingereza, Kiswahili na Kifini ameelezea jinsi programu zenyewe za kialgorithimu zinavyofanya kazi. Hata kama amemakinikia lugha ya Kiswahili katika SALAMA na

kwenye tafiti nyingine za kando alizofanya kuhusu teknolojia ya lugha, hajazamia kwa kina kuhusu nafasi ya teknolojia hii katika utafiti wa nyanja zote za Kiswahili kwa kama lugha kwa jumla. Katika makala yake kwa lugha ya Kiswahili, Hurskainen (2008) ameandikia kuhusu nafasi ya nadharia katika teknolojia ya lugha na nafasi ya teknolojia ya lugha katika utathmini wa nadharia mbalimbali za kiisimu na kihisabati. Kazi zote hizi zilijenga msingi wa utafiti wetu kwa kuwa Hurskainen ametalii nyanja zote za SALAMA kama muasisi. Hata hivyo, utafiti wetu haukumakinikia nadharia za kiisimu na kiisabati kwa undani au mbinu za kialgorithmu na kikompyuta ila umeangalia nafasi za mbinu hizo katika utafiti wa lugha ya Kiswahili.

Utafiti wa hivi majuzi kabisa ulitumia programu za kikompyuta katika uainishaji wa data maalumu za Kiswahili ni ule wa (Hans 2013) ambaye alitafitia mchango wa lahaja katika ukuzaji wa msamiati wa Kiswahili sanifu katika shahada yake ya uzamifu. Hans anasema kuwa aliweza kuainisha msamiati wa lugha kwa urahisi kutumia mbinu hii na akapendekeza mtazamo huo kuendelezwa ili teknolojia isaidie utafiti wa kiisimu wazo ambalo linalingana na moja ya haiypothesia za utafiti huu.

1.7. MISINGI YA NADHARIA YA UTAFITI

1.7.1 NADHARIA YA MSAMBAO WA UGUNDUZI

Nadharia hii pia herejelewa kama Nadharia ya mweneo wa Bidaa. Nadharia hii iliasisiwa na Everret Rogers ambaye ni profesa wa masuala ya mawasiliano na mabadiliko ya kijamii (social change) mnamo mwaka wa 1962 katika kitabu chake cha *Diffusion of Innovation* ila ina chimbuko lake katika mawazo ya mwana-elimu jamii Gabriel Tarde 1903 aliyejaribu kuonyesha jinsi ugunduzi na uvumbuzi wa aina tofauti unavyosambaa katika jamii mbalimbali. Katika miaka ya 1971, 1983, 1995 na 2003 Everret aliifanyia nadharia hii mabadiliko machache hatua kwa hatua katika matoleo hayo, la mwisho likiwa la tano katika kitabu chake cha 2003-*Diffusion of Innovation*. Kuanzia 1995 nadharia hii ilitumika katika tafiti mbalimbali, zilizohusu teknolojia ya mawasiliano na baadaye kumakinikia alichokiita teknolojia saidizi (*assistive technologies*) na jinsi zinavyopenya, kukubaliwa au kukataliwa na jinsi zinavyorahisisha kazi ya wanajamii husika. Kwa mfano, Everret aliitumia nadharia hii kuonyesha jinsi dunia ilivyounganishwa na mtandao wa intaneti miongoni mwa tafiti nyingine za teknolojia.

Everett (2003) anafasili msambao kama mchakato ambao kwao ugunduzi au teknolojia (anavitumia kama sinonimu) hupenya au huwasilishwa kupitia kwa mbinu au nyenzo mbalimbali katika muda fulani kwa wadau wa mfumo jamii fulani. Anafafanua ugunduzi kama wazo, desturi ya kufanya

jambo, njia mpya ya kufunza jambo, teknolojia mpya au ala na kadhalika inayochukuliwa na mwanajamii au kundi husika kama mpya. Nayo nyenzo anaiona kama chombo mfano tarakilishi, mbinu ya mawasiliano au mfumo wa kutekeleza jambo. Anaamini kuwa uamuzi wa kukubali kupokea ugunduzi huu kwa mwanajamii ni mchakato wa hatua tano (au sita) kulingana na alivyochagua mhusika;

- 1) Ufahamu – mhusika hujua au kufahamishwa kuhusu ugunduzi fulani na anakuwa na wazo jinsi unavyotenda kazi yake kiuamilifu.
- 2) Ushawishi– mwanajamii anajenga mielekeo chanya au hasi kuhusu ugunduzi huo.
- 3) Uamuzi– jinsi mtu huhusika katika harakati zinazomuelekeza katika kuchagua kukubali/kukumbatia au kuukataa yaani kuupuuza ugunduzi huo.
- 4) Utekelezaji/Ukubalifu – mtu kukubali kuweka ugunduzi huo katika matumizi yake ili umsaidie katika shughuli zake.
- 5) Uthibitishaji–Mhusika hutathmini matokeo ya uamuzi alioufanya juu ya ugunduzi fulani.
- 6) Kuacha kutumia-Baada ya kukumbatia teknolojia fulani au ugunduzi husika mwanajamii anaweza kuachana nayo kama haimfai tena au inapopitwa na wakati.

Katika ya hatua au mchakato huu, hatua ya kwanza, ya pili na ya nne ndizo tutakazozitumia katika utafiti huu. Hatua ya tatu hudhamiria kuonyesha jinsi teknolojia hii inavyotumiwa na watafiti wa Kiswahili. Teknolojia ya kompyuta imekuwa maisha ya kila msomi hasa watafiti. Tutachunguza jinsi wanavyobwia teknolojia hii au kama ni kiasi gani waliozikumbatia. Kipengele cha nne kingetumika katika katika kupima mielekeo na mitazamo ya watafiti wa Kiswahili kuhusu teknolojia hii ya lugha tkwa kugegemeza matumizi ya mbinu mbalimbali zilizosanidiwa tayari katika SALAMA. Hili halitofanyika katika utafiti wetu kwa kuwa liko nje ya upeo na mipaka ya utafiti huu. Tatalendekeza baadaye lishughulikiwe na mtafiti mwingine badaye.

Hatua ya ukubalifu na ambayo ni kipengele muhimu cha Nadharia ya Msambao wa ugunduzi kwa kuwa teknolojia husambazwa kupitia kwa lugha nayo lugha hufaidika kiutafiti kupitia kwa teknolojia. Kiswahili kimepiga hatua hiyo kwa jinsi kilivyotayarishwa kwa ajili ya matumizi katika tarakilishi na miradi kama SALAMA (Mabeya 2009).

Vigezo vitakavyotumika katika upimaji wa nadharia hii katika hatua tatu zilizotajwa hapo juu ni vitano kama asemavyo Rogers (2003: 207-238). Vigezo hivi ndivyo hupimia kasi ya ukubalifu na hasa huwa ndizo sifa mahsusi za ugunduzi ambazo husaidia na husukuma upenyaji au msambao wake

Vigezo au Sifa za Upenyaji wa ugunduzi

1. Manufaa/Faida/Istihaki rejelewa ya teknolojia mpya ni kiasi ambacho ugunduzi fulani unachukuliwa kuwa bora zaidi kuliko hali iliyotangulia. Hili hupimwa kwa nyongeza kiuchumi,

ufaafu, uokoaji muda, urahisishaji mchakato uliopo, uridhishaji mteja na kadhalika. Haitegemei kama kuna manufaa ya jumla kuhusi ugunduzi huo bora uwe umehisiwa kuwa nayo na mhusika. Kiasi cha mapendeleo au kuwaziwa kuwa na manufaa ndiyo masiala ya kasi ya ukubaliwaji wa ugunduzi. Kwa mujibu wa utafiti wetu, tutaangalia jinsi programu mbalimbali za kikompyuta katika mradi wa SALAMA zinavyofanya utafiti wa lugha ya Kiswahili kuwa rahisi na wa kuaminika kuliko mbinu za kiungamuzi na kimajaribio ambazo zikuwa zikitumiwa na watafiti wa lugha kwa muda mrefu. Nguzo hii ya istihaki au manufaa ya ugunduzi au teknolojia husika ni muhimu sana katika nadharia ya msambao wa ugunduzi kwa kuwa ndiyo huelekeza jinsi jamii husika inavyofaidi kwa kukumbatia ugunduzi mpya unaowajia kwa wakati huo na kuuhusisha katika shughuli zao. Tutatumia kigezo hiki kutathminia jinsi teknolojia ya lugha katika mradi wa SALAMA ilivyosaidia katika utafiti wa lugha kama njia ya kisayansi na vipi ilikuwa bora kuliko njia za kushughulikia lugha za kiungamuzi au kidhahania.

2. Ulinganaji/Ufaafu *ambao* ni jinsi ugunduzi au teknolojia huhisiwa kuwa inaenda sambamba na desturi, tamaduni, maadili, miiko, tajriba za awali na haja za watumizi watarajiwa. Wazo lisiloingiana na maadili ya kundi jamii fulani huwa vigumu kukubalika. Mfano ni dawa za kupanga uzazi kwa Wakatoliki na Waislamu (Rogers 2003). Tutahusisha kipengele hiki katika uchanganuzi wa data za utafiti wetu kwa kuangalia jinsi teknolojia ya lugha inavyoendana na wakati katika jamii hii ya karne ya 21 ambayo imeibukia kutegemea kompyuta kwa kila jambo na jinsi Kiswahili kinavyoweza kutafitika kutumia teknolojia.

3. Uwezekano wa kimatumizi ni jinsi wana-kundi jamii hao wanavyohisi ugunduzi fulani kuwa mgumu na utata kutumia. Dhana rahisi hukubalika kwa urahisi na dhana ngumu huhitaji mlengwa apate ujuzi mpya na uelewa mpya wa hali andalizi. Katika mfano anatoa mwasisi wa nadharia hii, wanakijiji wa Das Molinos kwa kuwa hawakuelewa dhana ya viini na magonjwa, hawakuweza kukumbatia dhana ya kuchemsha maji ya kunywa ili kujiepushia magonjwa. (Rogers 2003). Katika kipengele hiki tutaonyesha jinsi dhana za kufanyia utafiti lugha kikompyuta kama aina ya ugunduzi zilivyoweza kuwashirikisha wanalugha wengi kutoka janibu mbalimbali katika mradi wa SALAMA ambao ndio kifani cha utafiti wetu.

4. Ujaribishikaji ni uwezekano wa teknolojia mpya au ugunduzi fulani kutekelezwa kwa mawanda finyu kimajaribio. Je, teknolojia ya lugha iliyojaribishwa na kufanya kazi katika SALAMA inaweza kufanya kazi kwingineko? Mawazo yanayoweza kujaribika kwa kiasi kidogo hukubalika kwa haraka kuliko wazo kubwa la kuja kwa mkupuo mmoja. Majaribio ndiyo hulengwa na hatua za ushawishi na utekelezaji hapo juu. Watafiti wengi wa Kiswahili kama tulivyowanukuu katika udurusu wa yaliyoandikwa kuhusu mada yetu hii walirejelea mradi wa SALAMA kama kielelezo

cha utafiti wa lugha ya Kiswahili kilichofanikiwa na pia wengine kutumia kongoo ya HCS katika utafiti wao.

5) Ushuhudiaji au Uambukizanaji – hurejelea jinsi matumizi ya teknolojia mpya yanavyoweza kuonekana, kusikika au kushuhudiwa. Ushuhudiaji huu humaanisha kuwa wengine pia wanajaribu dhana mpya na hatimaye husaidia watu kupokea na kukubali dhana hizo.

Rogers, E(2003) anasema kuwa sifa moja ya kustaajabisha kuhusu nadharia ya Msambao wa Ugunduzi ni kuwa ukubalikaji wa ugunduzi fulani kwa kundi-jamii fulani hutegemea wanajamii wengine. Ndipo inasemekana kuwa grafu ya kuonyesha upenyaji wa ugunduzi hupanda kulia kwa umbo la herufi 'S'. Utegemeanaji huu umepingwa na wasomi wengine na hususan huelezewa kama udhaifu wa nadharia hii. Udhaifu mwingine ni chukulizi kwamba mbinu kama vyombo vya habari huathiri upenyaji wa ugunduzi, suala linalozua mjadala mrefu na utata mwingi kwa mujibu wa Melkote (1991) anayesema kuwa sio wote wanaoathiriwa na vyombo hivyo kwa kuwa wengine wana mawazo huria. Kama nilivyodokezea hapo awali, taasisi ya ACALAN imejiingiza katika mradi wa kuunda kongoo kubwa ya Kiswahili mapema mwaka ili hatimaye wasaidie utafiti wa Kiswahili kwa kuwa walishuhudia ufanisi wa mradi wa SALAMA. Ithibati ya dai hilo ni kuwa wanaongoza mradi huu yaani Prof.Kimani Njogu kutoka Kenya na Dkt.Suleiman Sewangi kutoka Tanzania wamenukuu mradi wa SALAMA katika pendekezo lao na kuridhia kuwa uliongoza njia katika shughuli hizi za teknolojia ya lugha katika utafiti wa Kiswahili.

Vipengele hivi kwa hivyo vitakuwa misingi yetu ya kutumia nadharia ya Msambao wa Ugunduzi katika uchanganuzi wa data katika utafiti huu sambamba na malengo yetu.

1.7.2 NADHARIA YA ISIMU KONGOO AU MTAZAMO WA KIKONGOO

Utafiti huu unahusu lugha katika mazingira ya kikompyuta. Tunahisi kuwa Nadharia ya Msambao wa Ugunduzi pekee haitaridhirisha mahitaji yote ya utafiti huu yakiwemo malengo, maswali ya utafiti pamoja na haipotesia. Hii ni kwa sababu, kando na suala la teknolojia kuwa ni ugunduzi unaosukuma utafiti na kurahisisha mchakato huo, kuna kipengele cha lugha yenyewe ambacho kinatafitiwa. Utafiti wa lugha katika kompyuta huegemea hazina ya lugha hiyo iliyohifadhiwa katika kompyuta ambayo ni neno au kipashio cha kileksika (Sinclair, 2004). Hazina hiyo, ambayo ni sehemu ya kiwakilishicha lugha katika matumizi yake halisi ndiyo huitwa kongoo. Nadharia inayotawala utafiti wa hazina ya lugha halisi katika mazingira ya kikompyuta huitwa Mtazamo wa Isimu Kongoo au Nadharia ya Isimu Kongoo au hata Mtazamo wa Kikongoo kama anavyouta Mahlberg, M. (2005). Nadharia hii imefidia upungufu wa kutoshirikisha uchunguzi wa 'maneno' ya lugha katika nadharia ya kwanza ambayo hasa yalimulikwa katika utafiti huu.

Nadharia ya Isimu Kongoo au Mtazamo wa Kikongoo katika uchunguzi wa lugha ni dhana ambayo ilianza kutumika katika karne ya 13 wakati wasomi wa Biblia wa Kikristo walipitia msahafu mzima na kuweka alama kwa mikono yao katika kila neno, mstari kwa mstari, ukurasa kwa ukurasa na kuhakikisha kuwa yameorodheshwa ili kila moja lirejelewe kwa maana yake halisi (O'keeffe.A na McCarthy. M.2010). Hii ilikuwa kazi ngumu sana ila bado iliendelezwa kwa kuwa umuhimu ulikuwa katika kutegemea lugha katika matumizi yake halisi. Kazi mahsusi inayorejelewa sana ni ile ya mwaka wa 1230 BK ambapo Kadinali Hugo ambaye alisaidiana na masufii au mawalii 500 wa Kidominika na wakafanikiwa kuunda orodha ya maneno ya mkusanyiko wa vitabu vya kidini ulioitwa Vulgate. Kazi kama hii iliyowachukua masufii 500 karibu mwaka mzima inaweza kufanywa siku hizi kutumia programu za kompyuta za kuunda upatanisho wa lugha (Computer Concordancing Programs) ndani ya sekunde moja au hata haraka zaidi (O'keeffe.A na McCarthy. M. 2010).

Kutoka miaka hiyo, kazi ya kuchanganua lugha katika kongoo kutumia mikono ilitamalaki na mnamo 1787 Becket aliunda uchanganuzi wa upatanisho wa matini za kazi za mwandishi mashuhuri William Shakespeare (A Concordance of Shakespeare).

Dhana hii iliwasaidia sana wasomi wa lugha kwa kuwa waliweza kupata hazina ya sehemu fulani ya lugha mahali pamoja na kuitumia kuchanganua lugha. Hii ndiyo dhana msingi hadi wa leo katika programu za kikompyuta za kuchanganua lugha. Jukumu muhimu kwa haya yote limekuwa ni kuorodhesha maneno, kuangalia ujirani wa maneno, kutathmini maneno yanatokea mara ngapi, kuangalia muktadha wa matumizi ya neno na kadhalika.

Kwa mujibu wa O'keeffe na McCarthy (2010), uundaji wa kongoo matini za kielektoniki (kwenye kompyuta) ulianza katika miaka ya 1950. Juhudi hizi zilianzishwa na wanaleksikografia na pia wanaisimu muundo wa kabla ya Chomsky. Wanaisimu-muundo wa Marekani wakiwemo Harris, Fries na Hill waliongoza mchakato wa uundaji kongoo kama njia ya kuibuka na lugha halisi kwa matumizi ya utafiti wa lugha (Leech 1992). Umuhimu wa kuwa na kongoo matini pia unaweza kungaliwa kwa mtazamo wa matumizi yake katika ufundishaji lugha ya kwanza kwa watoto kama inavyodhihirika katika kanzi-data ya CHILDES iliyoanza kuundwa miaka ya 1960 na kuzinduliwa rasmi mwaka wa 1984 katika kiungo cha tovuti www.childes.psy.cmu.edu. Miaka ya 1960 hadi 1970 ilipigia hatua fupi fupi kwa kutumia teknolojia ilijulikana kama *KWIC* -Key Word in Context (yaani kuweka neno kuu kwenye muktadha ufao ili uweze kutafuta machapisho maktabani na kadhalika kutumia kompyuta zenye uwezo mdogo sana. Maendeleo katika teknolojia ya kompyuta katika miaka ya 1980 n 1990 ndiyo yayolizaa wazo la kuwa na kongoo kama tuijuayo siku hizi yaani, data ya kutumiwa na wanaisimu na wataalamu wa lugha kwa utafiti wa lugha.

Wasomi wengi watajika ambao pia ni wanaisimu kama vile Leech, Biber, Johansson, Francis,

Hunston, Conrad, na McCarthy miongoni mwa wengine, bila shaka wamechangia kwa njia moja au nyingine katika uimarikaji wa taaluma ya Isimu kongoo ya zamani na ya siku hizi.

Licha ya hayo, wanaisimu wengi vile vile wanamchukulia John Sinclair kuwa mmoja wa wasisi wakuu kwa kuwa ni msomi mwenye msukumo na athari kubwa zaidi katika sura ya Isimu Kongoo ya siku hizi. Sinclair ndiye alisemekana kugundua kuwa neno lenyewe au ndani yake halibebi maana, ila maana hujitokeza panapokuwa na maneno kadhaa katika ruwaza au mpangilio fulani (Sinclair, 1991). Mrengo wa kuhusisha muundo na maana ilizaa haja ya kuwa na kongoo kubwa.

1.7.2.1 Nguzo Za Nadharia Ya Isimu Kongoo

John Sinclair kwa hivyo ni mtaalamu anayechukuliwa kama baba wa Isimu Kongoo au mwasisi wa mtazamo wa kikongoo katika uchunguzi wa lugha. Sinclair amechapisha na kuhariri vitabu zaidi ya 30 na makala zaidi ya 100 katika uwanja wa sarufi, msamiati, uchanganuzi wa diskosi, leksikografia, mtindo, ufundishaji lugha na bila shaka Isimu Kongoo. Katika miaka ya 1970 Sinclair alileta mapinduzi katika utafiti wa diskosi ya usemi na kufikia miaka ya 1980 aliunda mradi mkubwa zaidi ulimwenguni wa utafiti wa Kiingereza kuwahi kuonekana ambao ulizaa Kongoo ya maneno 400 milioni ya Kingereza iliyojulikana kama Bank of English Corpus na kuibuka na aina mbalimbali za makamusi, sarufi na marejeleo ya kufundishia lugha ambayo yalikuwa sehemu ya mradi wa COBUILD (Sinclair, 1991).

Kamusi ya kwanza ya mradi wa COBUILD ilipuuzwa na wachapisha vitabu wengi, ila baadaye ilikuja kuibuka kuwa mashuhuri na yenye athari kubwa kiasi cha kwamba kila kamusi na vitabu vya marejeleo hasa vya wanaojifundisha Kiingereza kama lugha ya pili, sasa vina misingi yao katika kongoo matini na vimeathiriwa na kanuni za ufafanuzi na utafiti uliosiswa na John Sinclair. Kazi maarufu zaidi za Sinclair ambazo ni msingi imara wa nadharia ya Isimu kongoo ambazo zililipa neno nafasi kuu katika nadharia hii, ni vitabu vyake *Corpus, Concordance and Collocation* (Sinclair 1991) na hatimaye *'Trust the Text'* yaani 'Amini Matini' (Sinclair 2004) ambacho kinafafanua nguzo kuu ya nadharia ya isimu kongoo kuwa ni maneno ya lugha katika matumizi yake na huo ndio anaouita ushahidi mkuu wa lugha. Katika miaka ya 1990 aliendelea kusukuma ugunduzi wa ufafanuzi hasa katika uchanganuzi wa kikompyuta wa mipangilio ya misamiati na sarufi na uchanganuzi wa diskosi na hapo ndipo alipoibuka na kipengele cha nadharia hii kuwa msamiati ni muhimu sana katika utafiti wa lugha.

Bado kuna mijadala mirefu na hata migogoro mingi tu pamoja na kutokubaliana iwapo Isimu Kongoo, kimsingi, ni taaluma au mbinu ya kiutafiti yaani (methodology) au inahitaji misingi yake ya kinadharia ya kujisimamia. Wanaoandama tapo la kuwa Isimu Kongoo au mitazamo ya kiutafiti inayoegemea kongoo matini katika ufafanuzi wa lugha ya Kingereza kwa mfano, wanadai kuwa

mbinu na zana mpya yaani programu za kikompyuta zinahitajika ili kuweza kuelezea hali ya matini-halisi. Programu au mbinu ambazo zimetumika katika kutafiti kwenye kongoo ni kama LEXA, MonoConc, MicroConcord, TACT, WordSmith, WordCruncher, Antconc3.0, na pia mbinu nyingine za kujiundia ambazo ni mahsusi kwa utafiti na kongoo husika kwa mfano vichambuzi vya kimofolojia kama SWATWOL na vifundisha lugha kama SWALEARN zilizotumiwa na Hurskainen kushughulikia kongoo ya HCS ya Kiswahili na kadhalika. Tayari mawazo ya misingi ya kinadharia yanayokumbatia zana hizo yanaanza kuibuka na kukita mizizi.

Mahlberg (2005) anaitetea kazi za Isimu Kongoo ambazo zina athari za kinadharia na zina madai ya kinadharia kama kazi za kina Teubert (1999), Hunston na Francis (2000), Sinclair (2004) na Hoey (2005).

Katika msingi wa madai yake Mahlberg, anatoa sifa kuu bainifu za Nadharia ya Isimu Kongoo kama ifuatavyo kwa muhtasari;

- 1) Lugha ni tukio la kijamii na maana inaonekana kama matumizi yake.
- 2) Kongoo matini ni lugha halisi na hivyo huhitaji zana au programu (za kikompyuta) kuelezea
- 3) Ruwaza au mipangilio mbalimbali ya matumizi ya lugha huonekana katika kongoo na ushahidi wa kongoo huonyesha kuwa muundo na maana vinahusiana. Maelezo ya kiisimu kongoo ya kufafanua lugha kutoka chini-kwenda-juu yanaupa msamiati kipa-umbele. (Mahlberg 2005).

Nguzo hizi mbili za kwanza za nadharia hii zinajumuisha mawazo na mantiki ya kimsingi kuwa, kinyume na yale ya Chomsky ya kuwa na umilisi wa lugha akilini tusiouona, lugha halisi ikihifadhiwa kwa ukamilifu wake na kuchanganuliwa kutumia mbinu za kikompyuta ambazo ni za kisayansi na hazipendelei, ndio ushahidi tosha na wa kuaminika wa kufanyia utafiti wa lugha katika nyanja yoyote.

Strazny, Philipp (Mh) 2005:245 naye anaongeza kuwa kuwa licha ya kuwa miundo ya lugha imekuwa ikitafitiwa kizamani au kimapokeo kwa kutumia mbinu ambazo si za kiujarabati na hata kutegemea ungamuzi wa mtafiti, matumizi mapana ya tafiti zenye msingi wa kongoo, zinazofafanua njia anuwai za matumizi ya lugha zifanywe katika miaka ya 1980 na 1990. Wanawaunga mkono Biber na wenzake (1996:115) wanaosema kuwa ...'ni katika eneo la matumizi ya lugha ambapo utafiti unaotumia mbinu za kikongoo umekuwa na athari kubwa'. Wanaisimu kongoo waliotangulia walimakinikia kutambua ni mara ngapi kuna utokeaji wa aina za maneno katika kongoo k.m nomino, vitenzi, vivumishi n.k. lakini uundaji wa programu na mbinu zenye uwezo zaidi zimewawezesha watafiti kutambua na kuchambua ruwaza za mhusiano changamano ya kileksika na

kisarufi katika lugha yoyote kwenye kongoo.

1.7.2.2 Udhaifu wa Nadharia ya Isimu Kongoo

Strazny, Philipp (mh) (2005: 246-247) anatahadharisha kuwa, licha ya uwezo mkubwa wa mbinu za kikongoo kushughulikia kwa njia anuwai matini za lugha katika hali yake halisi, kuna uwezekano wa kuwa na upungufu hasa kongoo isiposhughulikiwa kikamilifu inapoandaliwa na kupangwa vyema kabla kuitafitia kikompyuta. Nguvu za utafiti wa kiisimu kongoo zimo kwenye uchunguzi wa matumizi ya lugha, shughuli inayohitaji uchunguzi wa kisayansi na kiujarabati wa hazina kubwa ya kanzi-data ya matini asilia ya lugha. Kwa hivyo nafasi ya juhudi za binadamu lazima iwepo mwanzoni ndipo utafiti ufaulu.

Udhaifu wa Nadharia ya Kongoo unaweza kufikiriwa kwa mikabala ifuatayo (Strazny, Philipp (mh) 2005);

- 1) Uundaji wa kongoo huchukua muda mrefu sana pamoja na kualamisha aina za maneno na kushughulikia mategu, kugeuza usemi kuwa matini na sifa nyinginezo.
- 2) Aina ya data za kongoo zinazokusanywa huenda ikawa inahitajika matini kubwa katika seva na pia inahitaji kufanyiwa nyongeza za kisasa. Hii ni kazi nyingi mno.
- 3) Kutegemea kompyuta hubidi watafiti wazamie matini na kupuuza data za usemi yaani kongoo-sauti.
- 4) Mwisho, ni upungufu wa kifalsafa kwamba ni sampuli ya idadi fulani ya maneno ilitolewa kutoka kwa idadi kubwa mno isiyojulikana ya maneno. Ni bora kuwa na tahadhari katika baadhi ya ubashiri au mahitimisho tunayoibuka nayo.

Hata hivyo, licha ya uwezekano wa kuwa na udhaifu huu, ni vyema ifahamike kuwa utafiti wa kikongoo tofauti na mbinu nyingine nyingi za kushughulikia lugha unaweza kufanyia utafiti kila nyanja ya lugha katika matumizi yake ikiwemo sintaksia, mofolojia, uundaji kamusi, semantiki, fasihi, historia na mapisi ya lugha, isimu jamii kama masuala ya lahaja, tafsiri kadhalika. Aidha, Isimu kongoo inatumia programu na mbinu anuwai zinazoweza kushughulikia utafiti mpana na wa viwango vikubwa katika hali changamano za lugha ambayo ni shughuli muhimu sana kwa muda mfupi ajabu. Kongoo ndogo sana pia zimetathmini kamusi zinazoenziwa na kuzikuta kuwa na kasoro nyingi ajabu. (Hurskainen 2002, 2004, 2009)

1.8 HAIPOTHESIA ZA UTAFITI

- i) Mradi wa SALAMA ulitumia programu za kikompyuta katika Utafiti na Uainishaji wa Lugha ya Kiswahili.
- ii) Mbinu za Teknolojia ya Lugha zilizotumika katika Utafiti wa Kiswahili kwenye mazingira ya

Mradi wa SALAMA zilishughulikia lugha katika nyanja mbalimbali.

iii) Mbinu na programu za kikompyuta zilizotumika katika Mradi wa SALAMA katika Utafiti wa Lugha ya Kiswahili kwa kutumia Teknolojia ya Lugha zinarahisisha na kufaidi taaluma ya Kiswahili.

iv) Kongoo Matini ni hitaji kuu na kiungo muhimu sana katika utafiti wa kiteknolojia ya lugha ya Kiswahili.

1.9 MBINU ZA UTAFITI

Utafiti huu ulishirikisha mbinu mbalimbali za ukusanyaji wa data na pia uchanganuzi na uwasilishaji wa data hizo.

1.9.1 Ukusanyaji wa data

Kazi hii imetumia njia kadhaa katika ukusanyaji wa data yake. Data katika utafiti huu pia ni ya aina mbalimbali ili kukidhi malengo yaliyowekwa na kujibu maswali ya utafiti huu.

1.9.1.1 Aina za data

Utafiti huu umetumia data mseto yaani data za msingi na data za upili.

I) Data za msingi

Data za msingi katika utafiti huu pia zinaweza kuwekwa katika makundi mawili tofauti.

a) Hojaji

Hojaji zilitumika katika ukusanyaji data za utafiti huu kwa kuwapelekea wahusika kwa barua pepe baada ya kuwasiliana nao kwa njia ya simu na ya barua pepe, na kupata jibu kuwa wako radhi kushiriki kwenye utafiti huu. Tumetumia na hojaji za aina mbili ziliwalenga wahusika wa aina hizo mbili. Hojaji y kwanza ilikuwa ni kwa Prof Arvi Hurskainen ambaye ndiye mwasisi wa mradi wa SALAMA katika Chuo Kikuu cha Helsinki nchini Finland. Utafiti hii unahusu jinsi mradi huu wa SALAMA ulivyo kielelezo cha ufanisi katika matumizi ya teknolojia ya lugha katika utafiti wa Kiswahili. Uteuzi wa msomi huyu kama chanzo cha data kupitia hojaji umeongozwa na umuhimu wake, dhima yake na ujuzi wake kuhusu mradi ambao ndio kifani chetu katika utafiti huu. Mbinu za kikompyuta zinazotumika katika mradi huu na habari nyingine kuhusu mradi wa SALAMA ambazo hazipatikani kwingineko zinalengwa kupatikana kutoka kwake. Uteuzi wa Hurskainen, ambao ni wa kielelekezi aidha umeongozwa na matini ya kuwa tutaweza kupata habari kuhusu hali ya sasa ya SALAMA ikilinganishwa na ulivyokuwa mradi huu miaka 30 iliyopita ulipoanza. Kutokana na hojaji hii tumefahamu mbinu zaidi zilizoshirikishwa na pia mipango ya mustakabalini kuhusu mradi wa SALAMA kwa mujibu wa malengo ya utafiti huu. Pia msomi huyu ndiye mjenzi mkuu wa

Kongoo ya Kiswahili ya Helsinki (HCS) ambayo ni sehemu muhimu ya utafiti huu na moja ya malengo makuu ya utafiti huu na kupitia kwa hojaji hii tumefahamu kuwa kongoo ya HCS inazidi kuongezewa maneno.

Hojaji ya pili ilikuwa ni kwa wataalamu wawili wa teknolojia ya lugha walioshiriki katika utafiti na ujenzi wa mradi wa SALAMA pmoja na Prof Hurskainen alipowasimamia kama mwalimu wao katika kazi zao za uzamifu ambazo ziliegemea kwenye mradi huu. Hawa ni Dkt Suleiman Sewangi-Baraza la Kiswasili la Taifa- Tanzania na Dkt Wanjiku Nganga kutoka Chuo Kikuu cha Nairobi. Ushahidi kuwa walishiriki katika mradi huu unaonekana katika tovuti ya SALAMA kwani picha na maelezo kuwahusu wawili hawa pamoja na Prof Hurskainen zinapatikana katika ukurasa wa mwanzo wa tovuti ya SALAMA. Hojaji kwa wataalamu hawa ziliibuka na data kuhusu programu za SALAMA walizoshirikisha katika utafiti wao wa lugha ya Kiswahili na jinsi zilivyowafanya kazi kikompyuta. Hili ni lengo kuu la utafiti huu. Hata hivyo ni mmoja wao-Dkt Sewangi aliyeweza kujibu hojaji.

b) Kongoo-Matini

Sehemu mbalimbali za kongoo matini ya Kiswahili ya Helsinki (HCS) yenye maneno 12.5 milioni zilitumika kudhihirisha kuwa baadhi ya mifano inayopatikana katika kazi za wataalamu mbalimbali inatokana na kongoo. Kongoo hii hupatikana katika tovuti ya ⁶hifadhi ya Lugha ya Finland katika www.csc.fi baada ya kutuma maombi ya ruhusa ya kufanya utafiti.

Pia kongoo hii itatumiwa kujaribia maneno ya Kiswahili kutoka sajili na taaluma mbalimbali ili kutathmini ukamilifu wake na uwezo wake wa kufaa kuwakilisha lugha ya Kiswahili kikamilifu licha ya kuwa ili lilifanywa na watangulizi wangu ikiwa ni pamoja na kutathmini msamiati wa kompyuta (Hurskainen 2002).

11) Data za upili

Utafiti huu pia utakuwa kwa kiwango kikubwa ni wa data kutoka maktabani na kutoka kwenye tovuti mbalimbali kwa kiasi kikubwa. Data hizo zitakuwa katika makundi yafuatayo ili kuibua habari zifuatazo.

- a) Ripoti kumi na tisa 19 za kitaaluma ambazo pia ni makala yaliyoandaliwa kitaalamu na Arvi Hurskainen ili kuelezea na kutoa habari kuhusu jinsi nyanja mbalimbali za lugha ya

.....
⁶Kiungo cha kupata fomu ya kutuma maombi ya kuruhusiwa kutafitia au kufikia kongoo ya HCS ya SALAMA anazananyinginezo <http://www.csc.fi/english/customers/university/useraccounts/languagebank>.

⁷Kiungo cha kufikia ripoti 19 ambazo ni Makala ya kitaaluma ya mradi wa SALAMA aliyoundaa Hurskainen kwa miaka mingi: <http://www.njas.helsinki.fi/salama/technicalreports.html>

Kiswahili zilivyokuwa zinashughulikiwa na programu za SALAMA.

Data kutoka kwa ripoti hizi ndizo kwa kiasi kikubwa imekuwa mifano mahsusi na maelezo ya utekelezaji wa programu za kikompyuta zilizofafanuliwa zaidi katika sura ya pili. Data hii ilisaidia katika tathmini ya lengo na swali la kwanza, pili na tatu. Data hii inapatikana katika ukurasa wa kwanza wa tovuti ya SALAMA kwenye ⁷kiungo <http://www.nias.helsinki.fi/salama/technicalreports.html>

b) Kazi za kitaaluma za mwasisi wa SALAMA Arvi Hurskainen alizoandika kuhusu mradi wa SALAMA katika hatua zake zote. Mengi ni makala aliyowasilisha katika makongamano ya kimataifa ya teknolojia na kuchapisha katika jarida la Nordic Journal of African Studies na katika majarida mengine kutoka 1992 hadi wa leo.

c) Kazi za kiusomi za wataalamu wawili ambao pia walifanya kazi na SALAMA chini ya Hurskainen na kuwa wanafunzi wake wa shahada ya uzamifu yaani Sewangi(2001) na Ng'an'ga (2005) hasa tasnifu zao za Uzamifu na makala mengine yanayohusu SALAMA. Data kutoka kwa kazi hizi zitaonyesha jinsi programu za teknolojia ya lugha zinafaa katika utafiti wa lugha ya Kiswahili. Pia tulitathmini jinsi walivyotumia kongoo matini ya HCS kupata data zao.

d) Kazi za kiutafiti kuhusu teknolojia ya lugha katika utafiti wa Kiswahili zinazohusika na mradi wa SALAMA zilisomwa ili kupata mifano zaidi inayoonyesha jinsi programu hizo zilivyotumika katika nyanja za utafiti wa Kiswahili kimofolojia, kisemantiki, kifasihi, kisintaksia na kadhalika. Hizi ni kazi kama wataalamu kama Peter Waiganjo, de Shryver, Mwamzandi na de Gauw miongoni mwa wengine.

Data kutoka kazi hizi zimekuwa ni mifano iliyodondolewa na maelezo yaliyotolewa ili kuonyesha jinsi teknolojia ya lugha inavyotumika katika utafiti wa lugha ya Kiswahili sambamba na kipengele cha nadharia ya msambao wa ugunduzi kinachohimiza matumizi ya teknolojia kwa istihaki zaidi kuliko kushughulikia lugha kutumia njia za zamani za kiung'amuzi na kimapokeo. Pia data hizi zilichangia na kuelekea katika kujibu maswali ya utafiti huu sambamba na nguzo ya nadharia yetu ya pili ya Mtazamo wa Kikongoo ambayo huyapa matini ya lugha katika matumizi yake halisi, umuhimu na mashiko ya kiujarabati na kisayansi yanapotafitiwa kikompyuta.

e) Kazi za maktabani na mtandaoni za wataalamu mbalimbali ambazo zinazungumzia masuala ya teknolojia ya lugha katika ushughulikiaji wa lugha ya Kiswahili zilikuwa muhimu katika kupata uelewa zaidi wa eneo hili la utafiti. Kazi hizi pia zitatufaa katika utathmini wa mradi wa SALAMA tukiuangalia. Kazi hizi ni kama mradi wa kubadilisha matini kwa sauti wa Chuo Kikuu cha Nairobi wa LLSTI uliongozwa na Muchemi Gakuru, na kazi za Peter Waiganjo, De Gauw na De Shryver

kuhusu Kongoo Sambamba ya SAWA.

1.9.2 Uchanganuzi wa data

Katika kutimiza malengo ya utafiti huu, mtafiti aliipanga data iliyopatikana kutoka kwa hojaji na kutoka kwa kazi zilizosomwa kuambatana na maswali na malengo ya utafiti huu. Uchanganuzi wa data hii umetekelezwa kwa njia zifuatazo;

I) Kutoa ripoti ya hali halisi ya mradi wa SALAMA tangu ulipoanza mpaka hali yake ya sasa katika kila hatua ya utafiti kwa kutumia data hizi kimaelezo.

ii) Kuonyesha ithibati kwa kutoa mifano inayoafiki na ufafanuzi wa jinsi teknolojia ya lugha ilivyotumika na inavyotumika katika mradi wa SALAMA kufanyia utafiti Kiswahili.

Iii) Kutambua na Kubainisha maeneo ya lugha ya Kiswahili yaliyoweza kutafitiwa na yanayotafitiwa na programu za SALAMA na kutoa mifano mwafaka na maelezo. Nyanja hizi ni kama vile kimofolojia, kisemantiki, kifasihi, kisintaksia, uundaji wa kamusi, n.k.

iv) Kutoa mifano mahsusi ya utafiti wa lugha ya Kiswahili kuthibitisha kila eneo au nyanja ya utafiti wa lugha ya Kiswahili kutoka kwa ripoti za kitaaluma zinazoakisi hali halisi ya hatua za utafiti na ushughulikiwaji wa lugha ya Kiswahili kutumia programu za kikompyuta za SALAMA.

v) Kutathmini mifano hii na kubainisha manufaa ya kutumia mbinu hiyo ya teknolojia ya lugha ikilinganishwa na kutumia mbinu za kiungamuzi, kidhahania na kimapokeo kufanyia utafiti wa lugha sambamba na nadharia zilizopendekezwa.

1.9.3 Uwasilishaji wa data.

Data iliyokusanywa kutoka maktabani, mtandaoni na kutokana na hojaji na kuchanganuliwa katika utafiti kama tulivyoeleza hapo juu itawasilishwa kulingana na mahitaji ya malengo na maswali mbalimbali ya utafiti huu na kama habari mahsusi inavyohitajika katika sehemu husika katika kila hatua ya kazi hii.

Uwasilishaji wa data hii umekuwa ni katika maelezo, mifano, tathmini ya programu za SALAMA katika utafiti wa lugha ya Kiswahili na ubainishaji wa manufaa na istihaki za matumizi ya programu hizo.

Pia kumekuwa na majedwali, michoro, picha na viambatisho vya data mbalimbali kutoka kwa mifano, mbinu na vyanzo vilivyotajwa ili kujibu maswali ya utafiti.

Uwasilishaji pia umehusu kudhihirisha jinsi data za utafiti huu zilivyoafiki haipotesia za utafiti kwa misingi ya nadharia mbili za Msambao wa Ugunduzi na Mtazamo wa Isimu Kongoo ufaafu wake katika utafiti huu.

SURA YA PILI

PROGRAMU ZILIZOUJENGA MRADI WA SALAMA

2.0 Utangulizi

Sura hii imeangazia baadhi ya vipengele, mbinu au programu mbalimbali zinazotumia kompyuta ambazo zinaujenga mradi wa SALAMA. Sehemu hii imefafanua baadhi ya programu kuu au mbinu kuu za kiteknolojia ya lugha zilizodhihirisha hali ya kushughulikia lugha ya Kiswahili kwa kuichanganua na kuitafitia kutumia nyenzo zilizojengwa *kialgorithimu* katika mradi wa SALAMA. Aidha sehemu hii ni muhimu katika utafiti huu kwa kuwa imeweka wazi suala kuwa mradi wa SALAMA ni mmojawapo ya miradi ya mazingira ya kikompyuta uliopiga hatua kubwa katika kuchanganua lugha ya Kiswahili. Tumejadili kimsingi jinsi programu hizi zinavyofanya kazi katika utafiti wa Kiswahili ambalo ni moja kati ya yale malengo makuu ya utafiti huu. Kudhihirisha kuwa mazingira ya lugha kikompyuta yanayofasiriwa katika utafiti huu kama SALAMA, yalitumia programu ambazo zilichunguza lugha kikompyuta linakuwa ni jambo muhimu pia katika kutusaidia kujibu swali letu la kwanza na kupima haiyohusiana na swali hilo.

2.1 Historia fupi ya Mradi wa SALAMA

Kwa mujibu wa Huskainen (2004), mwenyewe aliporudi Helsinki kutoka Tanzania mnamo 1985, alishiriki katika kozi iliyoendeshwa na mwenzake Kimmo Koskenniemi kuhusu Mofolojia ya Tabaka Mbili, na akajipa changamoto ya kujaribu mfumo huo katika lugha ya Kiswahili kwa kuwa ni lugha yenye utata wa kimofolojia kama Kifini. Ilikuwa changamoto kubwa ikijulikana kuwa Kiswahili ni lugha yenye viambishi na minyambuliko mingi.

Hurskainen 1995 anakiri pia kuwa shughuli hii ya kuunda programu ya kuchanganua lugha ya Kiswahili kimofolojia ilitiliwa nguvu zaidi na utafiti wa DAHE-(Dar-Helsinki Swahili Archives) uliokuwa ukiendelea katika Taasisi ya Uchunguzi wa Kiswahili katika Chuo cha Dar es Salaam kutoka mwaka 1988-1992 uliohusisha matini zilizopatikana kutoka kwa kanda sauti na faharasa za Kiswahili pamoja na makamusi yaliyokuwa kwenye kompyuta.

Matini za Kiswahili ziliendelea kuongezewa katika kanzi hii kutoka kwa kongoo ya maneno iliyokuwa tayari imeanza kukusanywa na pia kutoka kwa maneno mapya, makala kutoka nakala za kurasa za vitabu vya Kiswahili, istilahi zilizoundwa na ³BAKITA- (Baraza la Kiswahili la Taifa) kutoka magazetini na hatimaye kufikia mwaka wa 1997 kutoka kwa intaneti, mchango ambao uliharakisha maendeleo ya SWATOL, na hivyo ya kuimarika kwa SALAMA kwa mujibu wa Hurskainen 2004.

SALAMA ilianza ikiwa kichanganuzi cha kimofolojia cha matini za Kiswahili. Mtaalamu Kimmo Koskenniemi ambaye alifanya kazi pamoja na Arvi Hurskainen katika Chuo cha Helsinki alikuwa ameandika tasnifu yake ya uzamifu miaka miwili kabla ya hapo kuhusu mada *Two-Level*

Morphology: A General Computational Model for Word-Form Recognition and Production, (Mofolojia ya Tabaka mbili: Mfumo wa kuunda maneno na kutambua aina ya maneno kutumia mbinu za kijumla za kikompyuta) Hurskainen (1999). Mfumo huu uliibuka baadaye kuwa ndio njia ya kawaida ya kuainisha lugha. Jaribio la kwanza la kuunda programu ya SWATWOL (Swahili Two-Level Morphology) lilifaulu katika kozi hii ambapo mnamo 1992 maelezo ya programu hii yalichapishwa (Hurskainen, 1992) na hivyo kutoa programu ambayo kwa kiasi kikubwa ingechanganua mofolojia ya Kiswahili kimsingi na iliweza kutambua na kuchanganua maneno 10,000 ya Kiswahili.

Tayari programu hii ya kushughulikia Kiswahili ingelinganishwa na mifumo bora ya wakati huo wa vichanganuzi vya kimofolojia vya lugha nyingine zile. Utafiti kuhusu vibadilisha hali za kimofolojia vya Muundo-ukomo wa ⁸*Nipe-nikupe* (tafsiri yetu ya *Finite-state transducers*) ulikuwa umeendelea kushika kasi kwa muda (Karttunen 1994; Pulman and Hepple 1993).

Kufikia mwaka wa 1993 Koskenniemi alikuwa tayari ameunda programu au kichanganuzi cha kikompyuta cha kuainisha lugha zenye mofolojia changamano kama Kifini.

Kichambuzi cha SWATWOL kwa hivyo kilizidi kuboreshwa na idadi yake ya maneno (msamiati) ikaongezwa kutumia mbinu mbalimbali. Orodha mbalimbali za maneno na makamusi yalikuwa msingi wa nyongeza hiyo ya maneno lakini pia iliendelea kupanuliwa na matini mbalimbali zilizojengwa katika mkusanyiko wa ⁹DAHE (Dar-Helsinki Archives of Swahili). Kwa hivyo mkusanyiko wa maneno katika SWATWOL uliendelea kuboreshwa kwa kuongezea maneno zaidi na kuonyesha uwezekano wa upungufu katika baadhi ya maneno hayo, Hurskainen (1999). Kwa mujibu wa mtaalamu huyu, inastaajabisha kuwa hakuna aliyepanga kuanzisha mradi wa SALAMA ila hatua moja ilizaa hatua ya pili hadi ikafikia kiwango cha kuwa chombo bainifu cha kushughulikia lugha ya Kiswahili.

⁸***Vibadilisha hali za kimofolojia vya Muundo-ukomo wa Nipe-nikupe*** (tafsiri yetu ya Finite-State Transducers)- ni mfumo wa kuainisha vipengele vya kisarufi vya lugha kwa kuangalia jinsi kila kiambajengo chake kutoka kushoto kinavyochangia matokeo ya zao zima la kisarufi kwa mfano neno au sentensi. Lipps (2011:18-20). {ni}:["+Anim+" "+Sg+" "+P1+"] {me}:" +Perf+" {zi}:" +C10+" p i k a. Hata hivyo haipo katika mawanda ya utafiti huu kufafanua kwa kina jinsi vipengele hivi vya kiisimu vinavyofanya kazi.

⁹DAHE (Dar-Helsinki Archives of Swahili)- Hazina ya mkusanyiko wa maneno ya Kiswahili yaliyokuwa yakikusanywa na wanafunzi katika Chuo Kikuu cha Dar es Salaam ambao ulizaa msamiati uliotumika na SWATWOL na kuongezewa baadaye katika kongoo-matini ya HSC-Helsinki Corpus of Swahili. Mradi huu ulishirikisha Chuo Cha Helsinki na Chuo Kikuu cha Dar es Salaam na uliongozwa na Prof Arvi Hurskainen-Mwanzilishi wa SALAMA. (Hurskainen 1992).

programu-changanuzi hiyo ya kikompyuta na pia kujenga mazingira mwafaka ya kuunda na kujaribia kanuni hizo. Katika kuunda na kujaribisha kanuni za uondoaji wa utata wa maana (*disambiguation rules*) iliwezekana kuimarisha mofolojia ya maneno hayo zaidi.

Hatua nyingine ya kujenga SALAMA ilifikiwa wakati programu ya SWACGP (Swahili Constraint Grammar Parser) yaani Kichanganuzi Sarufi Kikwazo ilipozinduliwa (Hurskainen 1996).

Kichanganuzi hiki kilitekeleza majukumu ya utatuzi wa utata wa maana katika matini zenye utata na pia kueleza lugha kwa misingi ya kisintaksia. Fred Karlsson katika miaka ya 1990 na 1995 katika kazi zake, alihusika sana na ujenzi wa misingi ya kinadharia ya kichanganuzi cha Sarufi Kikwazo kinachoweza kufanyishwa kazi naye mwenzake Pasi Tapanainen (1996), aliandika kufikia mwaka huu wa 1999, SALAMA ilikuwa imemakinikia kuunda sheria za lugha za kialgorithmu (kanuni za michakato ya kikompyuta) zinazohitajika katika utatuzi wa maana ya maneno (*word-sense disambiguation*) kwa kupangia shughuli zinazohitajika katika ubadilishaji wa matokeo ya uchanganuzi wa kisintaksia hadi kwenye lugha-unde ya InterLingua.

Hadi wa sasa, kupitia kwa ripoti zake za kitaaluma zilizoandaliwa na Hurskainen kutoka mwaka 1996-2014, SALAMA imepiga hatua kubwa katika uchanganuzi wa lugha ya Kiswahili na hivyo ilifikia kuwa na misemo na makundi ya maneno 10,000 zikiwemo semi 2,500 zilizo na sehemu moja kama kitenzi (Hurskainen 2008) kwa mfano –kula chumvi- na kadhalika. Pia imeweza kutumia kongoo yake katika kutathmini kamusi mbalimbali, kufanya tafsiri za mashine, kuunda kamusi, kukagua tahajia ya Kiswahili, kutafsiri kutoka Kiswahili kwenda Kiingereza, Kiingereza kwenda Kiswahili na Kiswahili kwenda kwa InterLingua. Mradi wa SALAMA pia umetumia programu zake katika kuunda mifumo ya kufundishia lugha ya Kiswahili kutumia kompyuta.

SALAMA iliwasilishwa kwa umma wa kimataifa kama mradi wa kikompyuta wa kuchanganua lugha kwa mara ya kwanza na Arvi Hurskainen katika kongamano la ishirini la kimataifa la Kiafrika (20th International Biennial Conference of the Africa Language Association of Southern Afrika-ALASA) mnamo Julai 5-9-1999 nchini Afrika Kusini. (Hurskainen, 2002).

2.2 Kichanganuzi cha Kimofolojia- SWATWOL

Hurskainen 1999, anaelezea maana ya SWATWOL- kuwa istilahi iliyoundwa kutokana na maneno ya Kiingereza Swahili Two-Level Parser ambayo ni programu ya kikompyuta ya kuchanganua lugha ya Kiswahili kimofolojia kama tulivyodokezea hapo awali. TWOL- ambalo ni ufupi wa Kiingereza cha Two-Level- ni dhana iliyoshughulikiwa na Kimmo Koskenniemi (1983) ambayo Hurskainen 2008 ametafsiri kama Mofolojia ya Tabaka Mbili.

Hurskanen 2004 anasema kuwa, TWOL huchukua kamusi ya kimofolojia ya Kiswahili yaani msamiti na kipashio au jalada la kanuni za lugha kama malighafi yake na kufanya uchanganuzi wa kimofolojia. Kufikia mwaka wa 2004 kulikuwa tayari kuna kiunda kanuni cha kikompyuta (*automatic rule compiler*) kwani awali ilibidi kanuni za kimofo-fonolojia ziundwe na watu kwa kutumia mifumo ya majedwali.

Naye Sewangi 2001, anafafanua kuwa SWATWOL imeundwa na vipengele vitatu yaani msamiati wa lugha ya Kiswahili, ambao kufikia mwaka 2004 ulikuwa takribani 10,000, kanuni za lugha na mfumo wa uchanganuzi wa maneno unaohusu uambishaji, unyambulishi, aina za maneno na habari nyingine ya lugha katika kiwango cha neno. Kwa mujibu wa Sewangi programu ya SWATWOL hutumia ishara mbalimbali zilizojikita katika kazi nyingi za kimofolojia za wataalamu kama (Ashton, 1944; Polome, 1967; Mukama, 1967; Khamisi, 1988) katika kutoa habari za maneno ya Kiswahili. Fauka ya ishara hizi kutoa uchanganuzi wa kimofolojia, pia zinafafanua sifa zakietmolojia, kiistilahi za kisemantiki, na kwa kiasi fulani uchanganuzi wa sifa za kisintaksia za maneno mahususi na pia kuonyesha mipaka ya kishazi katika maneno ya Kiswahili Sanifu. Licha ya uwezo huu wa SWATWOL na kuwa iko huru kuendelezwa kwa kujengewa ishara zaidi pindi haja inapozuka, Sewangi anatahadharisha kuwa licha ya kuwa programu ililenga kushughulikia mofolojia kamili ya Kiswahili Sanifu;

“...bado haiwezi kusemwa kuwa inawakilisha kikamilifu mofolojia ya Kiswahili...ila programu yenyewe ya uchanganuzi ni uakisi wa ishara na fasiri za ishara hizo kama zilivyotumiwa na aliyeunda kichanganuzi hiki cha kimofolojia...”
Sewangi 2001:

.....
¹⁰BAKITA- (Baraza la Kiswahili la Taifa) - Taasisi ya Lugha ambayo ni chombo cha Serikali ya Tanzania iliyo na majukumu ya kutafiti na kuidhinisha istilahi ikishiriakiana na vyuo na taasisi nyingine za Serikali. Katibu Mkuu wa BAKITA wa sasa ni Dkt.Suleiman Sewangi ambaye alifanya utafiti wa tasnifu yake kuhusu SALAMA na pia ni mmoja wa wasimamizi wa kongoo inayoundwa na ACALAN.

.....
...¹¹ilikuwa na kamusi yenye mofimu 25,000...kwa mujibu wa majibu ya Hurskainen 2015 katika hojaji ya utafiti huu, hali hii ni tofauti kwa sasa kwa kuwa anasema SWATWOL mpya ina uwezo Zaidi na ina vipashio vingi kuliko hii ya zamani

Programu ya SWATWOL ndiyo ilizaa mradi mzima wa SALAMA kama tulivyodokezea hapo awali. SWATWOL ilikuwa na kamusi yenye mofimu 25,000 na ingeweza kutambua maneno ya Kiswahili yapatayo 45,000 kwa mujibu wa Ng'ang'a (2005).

Hurskainen, 1999, alikuwa ametangulia kutoa sababu ya hali hii kuwa nomino nyingi huundwa kutokana na mzizi wa kitenzi moja kwa moja. Programu hii ilikuwa na aina mbalimbali za masamiati au kamusi ambazo zilitekeleza majukumu tofauti kama kujaribu msamiati wa kamusi za chapa, kuchanganua kisemantiki na pia kutambua etimolojia za maneno kwa hivyo matokeo ya SWATWOL yanaweza kubadilishwa kwa kugeuza programu kidogo au kuchagua msamiati ufaao jambo ambalo **Hurskainen, 1999** alishauri kuwa yote mawili kwa pamoja yanafaa zaidi. Pia anasema kuwa uundaji wa vitenzi kutokana na mizizi ya vitenzi ulifanywa kutumia mbinu mseto bila kuandika kila kitenzi kilichoundwa kama kipashio tofauti. (Hurskainen 1999:141)

SWATWOL ilifikia hatua ya kushughulikia viwango tofauti vya vipashio vya lugha mathalan tungo-funge kama vile nahau, methali na misemo kimofolojia pindi msamiati wa matumizi ya lugha ya kawaida ulivyozidi kuongezeka. Pia ingejihusisha na vitenzi vinavyorudiwa rudiwa, istilahi za sajili au nyanja maalum na maneno mengine ambayo si neno moja (Hurskainen 2004, 2004b, 2005:118). Maneno haya ambayo ni vikundi vya kipekee vya maneno (tungo-funge) vilihitaji kuangaliwa kwa njia tofauti. Anaongeza kuwa dhana hizi hushughulikiwa kwa kuainisha neno moja moja katika tungo husika, kuorodhesha kila kipengele cha tungo husika na kukifafanua kulingana na msamiati wa kimofolojia ulio kwenye kongoo na pia kutenga kila kipengele na kukiainisha kimofolojia kutumia kanuni zinazosheheni habari kamili ya kimofolojia (Hurskainen 2004b). Mbinu hii ya SWATWOL hutumika katika kushughulikia maneno ambayo tumeyaita tungo-funge (MWE) kama nahau yenye kitenzi na nomino k.m. *-piga moyo konde, Umoja wa Mataifa* au miundo ya kujirudia kama *pitapita, pigapiga* na kadhalika.

2.2.1 Sifa za kimuundo za SWATWOL

Kama tulivyotanguliza ni kuwa SWATWOL inachanganua maneno ya lugha ya Kiswahili katika viwango viwili yaani cha kimsamiati/kileksika na kiwango cha juu juu kama yanavyoandikwa na kama linavyotambulikana kimaana katika matumizi ya kawaida, yaani maana msingi. Hurskainen 2004 anafafanua hali hii kwamba kila herufi katika kiwango kimoja hulingana na herufi nyingine katika kiwango cha pili na kiango cha kileksika huwa juu. Anatoa mifano ifuatayo;

Mfano 1

- (a) Kiwango cha kileksika: #mUalimu#

Kiwango cha kimsingi: Omwalimu0 ambapo **u**-ya kileksika inatambulika kama **w**- kimsingi.

b) Kiwango cha kileksika: #NIumbu# #NIbuzi# #NIita#

Kiwango cha kimsingi: Onyumbu0 Om0buzi0 On0ta0

Data asilia kutoka kwa Hurskainen, 2004:367-368

Ukiangalia mifano hii miwili hasa huu wa b) utaona vile SWATWOL inavyoweza kufafanua hali mbalimbali za kimfofo-fonolojia za maneno ya Kiswahili. Tunaelewa kuwa haya yanawezekana kwa kuwa programu hii hupata habari za maneno kutoka kwa msamiati wa kileksika ulio katika kongoo iliyoainishwa kimofolojia. Hurskainen 2004, anaona hii ni njia moja mwafaka ya kuathmini usahihi wa kisarufi wa maneno fulani kwa kuamini kuwa kila neno la lugha husika lafaa kuwa kwenye mkusanyiko wa msamiati huu na kama halipo basi neno hilo si sahihi kisarufi.

Haya yote huwezekana kwa kutungiwa kanuni za tabaka mbili. Kwa mfano, hatua ya kupata neno la kimsingi hutokana na kanuni isemayo-U-itambulikane kama W katika neno la kimsingi iwapo kuna irabu inayofuatia kulia mathalan katika a) *Mwalimu* na *Mualimu* hapo juu. Sehemu ya b) itahitaji zaidi ya kanuni moja kwa sababu za kimuundo wa neno. Kinachosisitizwa hapa ni hali ya kutekelezeka wala si uwezekano wa kutekeleza kanuni hii kwa kuwa hata msamiati ulioainishwa kimofolojia unatosha.

2.2.2 Msamiati uliofafanuliwa katika mfumo wa tabaka-mbili

Maneno karibu yote ya Kiswahili yanayotumiwa katika mazungumzo ya kawaida wamehifadhiwa kama msamiati wa SWATWOL uliofafanuliwa kimofolojia. Mchakato huu huchanganua neno katika viwango viwili kwa kutumia muundo wa miti unaoshuka hatua kwa hatua ukiangalia hali za ngeli za nomino husika, wingi au umoja na kujikita katika mzizi wa neno hilo hadi katika hatua ambayo huwezi kushuka chini ya hapo. Neno likifika hapa linaweza kukubaliwa kuwa ni neno sahihi la lugha ya Kiswahili. Pia neno lisilopita mtandao huo wa msamiati huwa si sahihi kisarufi.

Si ajabu basi, SWATWOL imetumiwa katika kutathmini kamusi tano za Kiswahili kwa kuitumia kupitia ...

“...matini neno kwa neno na kujaribu kuona ikiwa neno ni Kiswahili. Ikiwa inaamua kwamba ni Kiswahili, basi inakubali na kutoa maelezo yote yanayotakiwa. Kwa mfano kwa nomino inatoa ngeli yake na kuonyesha ikiwa ni umoja au wingi. Kwa vitenzi huonyesha viambishi vyote vilivyopo pamoja na maelezo ya kila kiambishi, na pia lemma ya neno huonyeshwa...” (Hurskainen 2002:285)

Mfano huu hapa chini waonyesha kujenga msamiati kutumia maneno katika a) na b) hapo juu. (Mifano ya kina zaidi kuhusu kazi ya SWATWOL imeshughulikiwa katia Sura ya Nne).

Mfano 2

“... (2a) LEXICON Start

Noun “N ”;

(2b) LEXICON Noun

NounClassPref1/2;
 NounClassPref9/10;
 (2c) LEXICON NounClassPref1/2
 mU NounClassStem1/2 "= 1/2-SG ";
 wA NounClassStem1/2 "mU 1/2-PL ";
 (2d) LEXICON NounClassStem1/2
 alimu E "= { teacher } ";
 (2e) LEXICON NounClassPref9/10
 NI NounClassStem9/10 "= 9/10-SG";
 NI NounClassStem9/10 "= 9/10-PL";
 (2f) LEXICON NounClassStem9/10
 umbu E "= { hartebeest }";
 buzi E "= { goat }";
 ta E "= { wax }";
 taa E "= { lamp }";
 (2g) LEXICON E
 # #;
 END..." (Data asilia kutoka kwa Hurskainen, 2004:367-368)

2.2.3. Kanuni za Tabaka-mbili

Kanuni hizi ndizo hutawala jinsi SWATWOL inavyofanya kazi kwa kuhakikisha jinsi neno linavyotokea katika hali yake ya kimsingi ni sahihi. Hurskainen, 2004, ametumia mfano wa Ngeli ya 9/10 kuonyesha jinsi ya kuweka msamiati katika hali sahili hata kama neno la kimsingi linabadilika badilika.

(3) "Kutotamka U itokeapo kabla ya vokali"

U:w <=> m _ :V ; kama katika Mualimu na Mwalimu hapo juu

Kanuni hii ya (3) inasema kuwa: *badilisha neno kwenye leksika liwe na U kwa w ya neno la kimsingi w iwapo na iwapo tu kwenye upande wa kushoto kuna m (inayotambuliwa katika viwango vyote viwili kama m), na kwenye upande wa kulia kuna vokali ya neno la kimsingi...*Hurskainen 2004. Viambishi vya ngeli ya 9/10 ni vigumu zaidi kushughulikiwa kwa kuwa huhitaji kanuni nyingi (Hurskainen 2004). Kuingilia maelezo ya ndani ya kanuni hizi yatakuwa nje ya mawanda na malengo ya utafiti huu.

Kwa hivyo, SWATWOL kwa mujibu wa Ng'ang'a, 2005 inahusisha mfumo wa kuainisha vipengele vya kiisimu vya maneno kwa kufafanua sifa zake kama vile za kimofolojia ambapo vipengele kama aina za maneno, mnyambuliko wake, viambishi vyake, sintaksia, etimolojia, sifa fulani za kisemantiki au maana, na pia nyanja au sajili ya kimatumizi kwa mfano fasihi, biashara, siasa, afya na kadhalika. Msamiati au kongoo yake hueleza mofimu mahsusi ambazo zinaweza kuchanganuliwa na SWATWOL. Baadhi ya majukumu haya ya SWATWOL yamefafanuliwa zaidi katika sura ya nne kutumia data iliyochanganuliwa na programu hii katika ripoti za SALAMA na

katika kazi za Hurskainen na wanafunzi wake Sewangi 2001 na Ng'ang'a 2005 waliofanya kazi na SALAMA. Naye Kamau, 2005 anaona kuwa programu hii ya SWATWOL hutumika katika kuhariri matini za Kiswahili kwa kusahihisha makosa ya tahajia na ya kisarufi katika matini, wazo ambalo lina mashiko tukizingatia mawazo ya (Hurskainen 1992, 1999 na 2004) kuwa SWATWOL inatumia kanuni za kikompyuta zilizojengwa kutumia nadharia za kiisimu.

SWATWOL imetumika pia katika utathmini wa kamusi tano za Kiswahili ambazo tayari ziko katika chapa (Hurskainen 1994, 2002, 2004a) na kwa kuwa imechanganua msamiati wa lugha ya Kiswahili kimofolojia na msamiati wenyewe ni shirikishi kuliko kamusi nyingi, basi ilikuwa ni rahisi sana kutambua upungufu wa kamusi hizi (Hurskainen, 2004) kwa kuziakisi kwenye msamiati wa SWATWOL kikompyuta. Zaidi ya haya, SWATWOL pia inaondoa utata wa maana ya maneno kimsingi kwa kuyaeleza katika miktadha mbalimbali na kutoa uainisho wa ziada. Katika kuondoa utata huu ndipo kipengele kifuatacho cha SWACGP katika SALAMA kilizaliwa.

2.3 Kitatuzi cha utata wa maana na kichambuzi msingi cha Sintaksi- SWACGP

Kwa ukamilifu wake SWACGP ni akronimu ya Kiingereza inayosimamia Swahili Constraint Grammar Parser (SWACGP) au Kichanganuzi cha ¹²Sarufi Kikwazo ya Kiswahili. Inaitwa Sarufi Kikwazo kwa sababu inatumia kanuni zinazoondoa au kuteua- baadhi ya kipengele tata au vinavyohitajika kwenye tungo za Kiswahili na kuviondoa hadi pale mchakato huu utabakia na maana iliyokusudiwa kwa kuangalia muktadha.

Kazi ya programu hii ya SALAMA ni kuchanganua sintaksia ya maneno ya Kiswahili na kuondoa utata wa maana ya maneno yaliyopitia kwenye programu ya SWATWOL. Programu hii huwa na faili yenye orodha ya sheria za kisintaksia za lugha ya Kiswahili. Baada ya SWATWOL kutumika katika uainishaji wa maneno ya Kiswahili kimfo-fonolojia, baadhi ya maneno yanakuwa na utata wa kimaana kwa sababu kadhaa. Neno linaweza kuwa na utata wa kimaana linapotumika kama neno kujisimamia kwa sababu ya kuwa limeandikwa sawa ila lina maana mbili au zaidi, kwa mfano –mbuzi- ambalo linaweza kuwa na maana ya mnyama na –mbuzi- lenye maana ya kifaa cha kukuna nazi. Pia neno linaweza kuwa katika umoja na wingi kwa mfano katika ngeli ya 9/10 na liwe halibadiliki. (Hurskainen, 2004). Hapa ndipo programu ya SWACGP huhitajika.

Kichanganuzi cha ¹² **Sarufi Kikwazo ya Kiswahili**, au (SWACGP) Constraint Grammar Parser CG-2 kilichoundwa na (Tapanainen 1996, 1999) ni programu ya SALAMA ya kuandika kanuni za upunguzaji wa utata wa maana ya tungo za kisarufi kwa kufanya mambo mawili inavyohitajika (a) kuondolea mbali neno, mofimu au maneno au maana fasiriwa isiyofaa katika muktadha fulani (Hii ndiyo sababu ya kuitwa Sarufi Kikwazo) na (b) kuteua au kuchukua vipashio kinachofaa katika ufasiri wa maana lengwa (**Hurskainen 2004:371**).

Pia utata katika Kiswahili aghalabu hutokea kwenye baadhi ya mofimu za kihomografu za nomino za ngeli fulani. Kwa mujibu wa Hurskainen 2004, hali ya utata wa maana huwa inajidhihiri sana katika vitenzi, ambavyo mara nyingi huwa na mofimu mbili au tatu kama zile zenye utata wa aina mbili au tatu.

Kutokana na hali hii neno linaweza kuwa na utata likiangaliwa kivyake lakini linapowekwa kwenye muktadha, suluhu ya utatuzi wa utata inaweza kutekelezwa kwa sababu mfumo wa ngeli hutokea katika makundi manne ya mofimu ambayo ni makundi tofauti ya nomino, vivumishi, viwakilishi na mofimu za zinazoambatana na vitenzi. Kwa hivyo hata utata utokeapo kwa moja ya seti hizi za mofimu;

“... unapochunguza kirai chote, kwa mfano kirai tenzi au kirai nomino na kadhalika, utata unaweza kutatuliwa...kichanganuzi cha Sarufi Kikwazo kilitumika katika kuunda kanuni nyingi za utatuzi wa utata moja ikiwa ni kufuta maana ambayo haifai kwa muktadha huo(ndipo ikaitwa sarufi kikwazo)...au kueua ufaao...” (Hurskainen 2004:371).

Kufikia mwaka wa 2005, faili hii ya matini ilikuwa na orodha ya sheria za kisintaksia takribani elfu moja na mia mbili yaani 1, 200 (Ng’anga, 2005). Ilikuwa imebainika tayari kuwa takriban 50% ya miundo ya maneno ya Kiswahili yana utata wa aina mbili na kwa kuyashughulikia maneno haya katika SWACGP utata unapungua hadi 8% katika orodha ya maneno yaliyo katika kazi za kubuni, na asilimia 5% katika matini ya magazeti ya Kiswahili (Hurskainen 1996: 572). Kwa wastani ni takriban 94% ya maneno yenye utata huwa yametatuliwa kwa ufanisi mkubwa.

Matokeo ya uchanganuzi wa kimofolojia wa SWATWOL na Uondoaji utata wa SWACGP katika kiwango hiki ndiyo hasa hutumika na programu mbalimbali za kikompyuta kama vile vikagua-tahajia na ¹³vitenganisha maneno (tokenisers) (Ng’ang’a- 2005:38).

Uelewa wa sarufi ya lugha husika ni muhimu sana kwa mtafiti hasa wakati wa kuunda kanuni za sarufi Kikwazo na uchambuzi wa kisintaksia kama uliotumika katika SWACGP kama tulivyoshuhudia katika sura ya nne.

Hii ni kwa sababu bila kuainisha lugha kisintaksia, uhusiano wa viambishi vya kiima na yambwa katika sentensi zenye vitenzi elekezi vinaweza kuleta utata wa maana.

Ni kutokana na sababu hizi ndipo haja ya kuunda kanuni za kutegemeka za utatuzi wa utata na kuzipa kipaumbele kwenye hatua za mwanzo za vikwazo (*constraints*) na kuzipanga kulingana na utegemekaji wake (Hurskainen 2004:374).

¹³Vitenganisha maneno (*tokenizer*)- ni programu ya kompyuta inayochukua tungo-funge za lugha au sentensi na kuyaorodhesha maneno yake neno moja kwa neno moja kwa ili yaweze kuelezewa yakiwa vipashio huru kila moja. Mbinu hii pia hutumika kuhesabu maneno.

Aidha, ni muhimu ifahamike kuwa uchambuzi wa kisintaksia unaotekelezwa na SWACGP ni wa kimsingi mno. Kwa mfano, tunaweza tu kufahamishwa ni upande upi kitenzi kikuu katika sentensi kinatokea katika kiima bila kutujaza kuhusu pahali penyewe pa kitenzi hicho katika sentensi. Kwa mfano, Hurskaninen 2004:376 anatoa mfano wa neno ambalo ni kiima na kuwa upande wake wa kulia kutakuwa na kitenzi kikuu-kielekezi ambacho pia kitakuwa na yambwa upande wa kulia katika kanuni ya uainisho ya @SUBJ> kwa kiima, @OBJ< kwa yambwa na kwa kitenzi @FMAINVtr+OBJ> kama katika sentensi *Mtoto analima shamba*. Hatufahamishwi kwa kanuni mahsusi hasa vinaunganisha wapi vipashio hivi.

Ubora wa kichambuzi cha SWACGP ni kuwa huwa ni programu iliyojikita katika kiwango cha sentensi au tungo-funge za maneno na hivyo inaweza kusemekana kuwa inashughulikia lugha ya Kiswahili kwa kiwango ambacho kinakaribia sana lugha katika matumizi yake asilia. Kwa hivyo, suala la kutambua matumizi yanayofaa ya neno fulani katika tungo hizi kwa muktadha fulani, ambao umetiliwa msingi na uchanganuzi wa kina wa kimofolojia kutumia SWATWOL, ndiyo hasa kipengele kikuu cha ufanisi wa kanuni za SWACGP kama asemavyo **Sewangi, 2001:55**.

Dhima hizi za SWACGP za utatuzi wa utata wa maana na uchanganuzi wa kisintaksia wa kijuujuu, zitaangaziwa kwa kina katika Sura ya Nne tumetathmini nafasi ya programu hii katika utafiti wa lugha ya Kiswahili.

2.4 Kitatuzi-Kibahatishi cha Kuondoa Utata wa Matini za Kiswahili cha SWA–GUESS

Ng'ang'a (2005:37) anasema kuwa programu ya SWACGP inaweza kutambua maana halisi ya maneno au kuondoa utata huo kwa karibu 94% ila akaongezea kuwa viwango hivyo hivyo vinaweza kuboreshwa zaidi na hata kufikia 97% tunapotumia programu nyingine iitwayo SWA-GUESS kwa aina za utata zilizobakia kwa kuwa inatumia kanuni za kujaribisha katika utatuzi wake wa utata wa maana za maneno.

SWA-GUESS ni mseto wa neno Swahili na *Guessing* yaani kubahatisha au kudhania au kukisia. Hii ni programu ya SALAMA iliyoungwa kwa kutumia programu ya kikompyuta ya *Flex* ambao ni mfumo wa zana za kampuni ya Unix ambayo hutumia kanuni za lugha zilizotungwa na binadamu ili kuondoa utata wa vipashio vya Kiswahili vyenye utata ambao haukuweza kuondolewa na kivikwazo na programu ya SWACGP pale ambapo haingewezekana kuandika kanuni za Sarufi kikwazo (Hurskainen 1999). Licha ya nyongeza hii katika jitihada za uondoaji wa utata wa maana ya maneno katika lugha yoyote Hurskainen ametahadharisha kuwa tusitarajie kuondoa utata 100% kwa kuwa kuna hata baadhi ya maneno ambayo hayatolewi maana halisi hata na binadamu kwa kutumia bongo lake. Hatua ya SWA-GUESS huzaa hali zifuatazo kama matokeo ya mchakato huu;

muundo nje wa maneno- ukihusisha alama za uakifishaji pamoja na alama za ziada kama vile za lafudhi kali, kareti, sedila, umlati, lafudhi hafifu, makroni, *tilde* na kadhalika; mzizi wa neno (Lemma); uainishaji wa aina ya neno; habari kamili za kimofolojia; majukumu ya kisarufi ya vitenzi; uainishaji wa kisemantiki, habari za kietmolojia za maneno na faharasa za maneno katika Kingereza (Hurskainen, 1999).

2.4.1 Kanuni za kutambua maana halisi kwa Ubahatishaji wa kujaribujaribu

Mbinu hii ya SALAMA si ya kutegemewa sana ila inasadia katika kufidia udhaifu au kukosekana katika msamiati wa kimofolojia wa lugha husika au kosa la tahajia. Kwa mujibu wa Hurskainen (2004:374) maneno halisi ya lugha yanaweza kuwa yamekosekana katika msamiati ulioainishwa na vichanganuzi vya kimofolojia k.m SWATWOL au yawe na kosa la tahajia au uendelezaji wa neno na hivyo kuhitaji kushughulikiwa kando kivyake. Anatoa mfano wa jinsi sifa za nomino za Kiswahili ambazo zimeendelezwa visivyo zinavyoweza kubahatishwa kwa misingi ya viambishi ngeli. Sifa kama hizi za aina ya neno kwa hivyo zinaweza kusaidia katika uundaji wa kanuni za kubahatisha. Hali nyingine inayoweza kuundiwa kanuni kama hizi katika Kiswahili, kulingana na mtaalamu huyu, ni pale ambapo neno linaloanza na herufi kubwa linaweza kufikiriwa kuwa nomino maalum iwapo limetokea mahali tofauti sentensini na si katika mwanzo wa sentensi na kadhalika.

2.4.2 Kubahatisha kwa Misingi ya Uwezekano

Hurskainen (2004:374) anapendekeza katika SALAMA kuwa iwapo mbinu iliyofafanuliwa hapo juu itashindwa kufanya kazi, jaribio la mwisho la kupata maana halisi ya neno la lugha ni kubahatisha kutumia uwezekano hasa tunapolenga kueleza neno kisemantiki licha ya kuwa ni mbinu ambayo haiwezi kutegemewa sana. Anaeleza mbinu kadhaa za ubahatishaji huu ukiwa ni pamoja na kutumia kongoo-matini kubwa na kuhesabu mara ambazo neno limetokea katika ujirani wa maneno mengine au jinsi yalivyotokea pamoja maneno mawili au zaidi. Habari hizi zinaweza kuwekwa kwenye kanzi-data kama WordNet na hatimaye kutumika na miundo ya zana za lugha ambazo (Hurskainen 2005:122) anaziita Ramani inayojizatiti Yenyewe (*SOM-Self Organizing Map*). Zana kana hizi hupanga maneno kulingana na ujirani wake (*co-occurrence/co-location*) kwa mujibu Ng'ang'a (2003, 2005) mwanafunzi wa Hurskainen aliyefanya kazi na kongoo ya SALAMA katika kufanyia utafiti uondoaji wa utata wa maana ya maneno ya Kiswahili, na hata kuonyesha uhusiano huo ni wa viwango gani. Hurskainen (2005) anasema kuwa data kama hizi kwa mfano zinaweza kutumika katika programu kama SWATWOL na kusaidia uandaaji wa kanuni za kuondoa utata wa kisemantiki kwa kuwa ukadiriaji wa kibinadamu unaweza kusababisha makosa mengi tu.

2.5 Kichanganuzi cha kina cha kisintaksia cha SWAFGP

Programu ya SALAMA ya SWAFGP ni hatua mbele katika uchanganuzi wa Sintaksi ya lugha ya Kiswahili kwa kuwa inazamia uchambuzi wa kisintaksia kwa kina zaidi kuliko vile SWACGP ilivyoweza kufanya. SWAFDP katika ukamilifu wake ni Swahili Functional Dependency Parser yaani Kichambuzi cha Sarufi Amilifu Tegemezi ya Kiswahili na inahusika na uchanganuzi wa kisintaksia ulio kamilifu na wa kina ili kusuluhisha changamoto zilizopatikana katika programu za awali kama SWACGP kwa kuwa iliundwa ikiwa na uwezo wa kimuundo katika utekelezaji wake. (Hurskainen 1999:142) ambaye anakiri kuwa hata inatumika kuandaa vielelezo vya miti ambapo utegemezi wa kila neno hufahamika, anasema pia kuwa SWAFDP (Kichambuzi cha Sarufi Amilifu Tegemezi ya Kiswahili) inaweza kutumika sana na ni muhimu katika ujenzi wa programu za lugha za tafsiri-mashine. Kwa hivyo, ufafanuzi wa kila neno hutegemea uamilifu wake na linavyotegemeana na maneno mengine hasa lile ambalo ni kichwa cha tungo husika na umbali ya maneno hayo na kichwa.

SWAFGP ilijengwa ilianzishwa na kujengwa na (Järvinen and Tapanainen 1997) kama FGP na ilishughulikia Kiingereza na Kifini kabla Hurskainen hajaipa uwezo wa kuchambua Kiswahili. Kwa mujibu wa Hurskainen, 2004 kichanganuzi hiki huwa kina kanuni ambazo huchunguza eneo pana la tungo na hata kutambua umbali wa neno hadi jingine na mahusiano ya vijenzi vya sentensi hiyo yaani kichwa na maneno mengineyo. Baada ya kutenganisha na kuorodhesha maneno yote katika sentensi, kila neno moja hupewa nambari yake. Kutokana na hali hii mfumo wa SWAFGP hatimaye hutumia kanuni hizi kuunda kielelezo cha miti ambapo mtandao wote wa vitegemezi huonyeshwa na hivyo mfumo huu kujulikana kama Kichanganuzi cha kina cha kisintaksia au Kichanganuzi Kamili cha kisintaksia (Hurskainen, 2004:388). Mfumo wa SWAFGP unachota nadharia ya utendakazi wake kutoka na sifa maalum zilizoasisiwa na Järvinen and Tapanainen (1997: 5) ambazo husawiri sentensi kuwa na kichwa, aghalabu kichwa hiki huwa ni vitenzi ukomo, na maneno mengine sentensini kuhusishwa moja kwa moja au kwa njia nyingine. Vipashio vinane ambavyo ni sifa za msingi wa FGP zimejadiliwa katika sura ya nne kwa kina.

Natija kuu ya kutumia SWAFGP kama kichambuzi cha sarufi ya lugha ni kuwa muundo wa kisintaksia wa lugha hii unafafanuliwa kikamilifu na pia uwezekano wa kutumia viainishi hivi vilivyoelezea uamilifu wa kisintaksia wa maneno mbalimbali katika kutambua maana halisi ya maneno (*semantic disambiguation*) au kuondosha utata wa kisemantiki na pia kuainisha maneno katika tungo-funge kama methali na nahau (Hurskainen, 2004:376-377).

2.6 Kitafsiri-Oto cha Kiswahili kwenda InterLingua na Mifumo mingine ya Tafsiri-Oto

Programu hii ya SALAMA inatambulikana kama SWA-TO-InterLingua kwa Kiingereza kumaanisha kuwa ni programu ya tafsiri-mashine inayotumika kufasiri lugha ya Kiswahili hadi lugha-unde ya ¹⁴InterLingua na pia kwenda lugha nyinginezo. Programu hii hutumia ujuzi wa kimofolojia, kisintaksia na kisemantiki wa lugha ya Kiswahili ili matini ya lugha hii itafsiriwe hadi lugha nyingine. Kwa mujibu wa Hurskainen (2008:42), licha ya kuwa InterLingua ni lugha ya kinadharia na bado haipo, anaiona kama inayosimamia muundo wa lugha na kuunganisha kila neno katika maana za kisemantiki na kanuni za kuongoza utumiaji wa maneno hayo. Anaona faida ya lugha hii ikiwa hali yake ya uhoromo (kuwa katikati) na hivyo kupunguza kazi ya kuunda mbinu za lugha za watu. Kila lugha kwa hivyo ingekuwa lugha chanzi na lugha lengwa hasa pale lugha zinatofautiana. Mfano

“...Lugha A > InterLingua > Lugha B
Lugha B > InterLingua > Lugha A
Lugha C > InterLingua > Lugha D
Lugha D > InterLingua > Lugha C ...” (Hurskainen, 2008:43)

Kwa hivyo, Hurskainen (1999:142) anaeleza programuya SWA-TO-InterLingua kama kibadilisha sarufi kinachogeuzwa matokeo ya SWAGUESS na kuleta muundo unaotarajiwa na InterLingua. Programu ya SWA-TO-InterLingua hutumia habari za kimofolojia, kisintaksia na kisemantiki za lugha chanzi na kufasiri matini kwa lugha unde ya InterLingua inayotumia visawe vya kisemantiki vya lugha chanzi hiyo na Kiingereza kama iliyoonyeshwa hapo juu huku uanishaji wa aina nyingine ukibakia ulivyo. Mabadiliko haya hufanywa kupitia kwa kanuni mbili:

2.7.1 Kanuni za kubadilisha uainishaji wa maneno

Kanuni hizi hizi hubadilisha viainishi vya lugha chanzi na kuwa katika hali inayohitajika na InterLingua. Viainishi vya InterLingua ni sawa na vya Kingereza vya kimofolojia na kisintaksia kwa mujibu wa **Hurskainen 1999** na hivyo Kiswahili na InterLingua itatumia kanuni sawa katika ubadilishaji wa kanuni wa viainishi kwa kuwa lugha zote mbili zina mpangilio wa **SVO** licha ya kuwa mara nyingine Kiswahili huchukua pia **SOV** wakati yambwa ni kiwakilishi, kwa viambishi awali na pia katika mpangilio wa maneno katika sentensi za maswali na sentensi za hali ya kutendewa.

¹⁴**InterLingua**- Hii ni lugha unde iliyopangiwa kuunganisha ulimwengu mzima. Kwa mujibu wa **Hurskainen (2008:42)**, anasema licha ya kuwa **InterLingua** ni lugha ya kinadharia na bado haipo, anaiona kama inayosimamia muundo wa lugha na kuunganisha kila neno katika maana za kisemantiki na kanuni za kuongoza utumiaji wa maneno hayo. Ametumia daraja horomo ya InterLingua katika utathmini wa michakato ya Tafsiri-oto kwa Kiswahili au kutoka kwa Kiswahili

2.7.2 Kanuni za kupanga maneno-upya

Hurskainen 1999, anaeleza kuwa kanuni hizi hupanga upya maneno ya lugha chanzi katika utaratibu unaohitajika na InterLingua. Kwa mfano, iwapo mpangilio wa maneno katika lugha chanzi ni SOV basi maneno katika sentensi husika yatapangwa upya na kuwa katika SVO. Vigezo vinavyohusika ili kutekeleza kanuni hizi kutajadiliwa kwa kina katika Sura ya nne. Kwa mfano kuna NOUN+POSS+DEM+ADJ+NUM (k.m. viti vyangu hivi vizuri viwili), mfumo utatumia mbinu inayoshughulikia kirai–nomino kirefu zaidi na kutokea na mpangilio wa DEM+POSS+NUM+ADJ+NOUN (ambao katika InterLingua utakuwa: **(these my two good chairs)**).

Pia mbinu za SALAMA zimetumika pakubwa katika tafsiri-oto ya Kiswahili kwenda Kiingereza na kinyume chake zikishirikisha mbinu kadha kama uchanganuzi wa kimofolojia, uondoaji utata wa kisemantiki na maana ya maneno kwa kutumia muktadha wa matumizi katika kongoo, kutumia uafafanuzi wa miundo ya kisintaksia ambapo tafsiri kamilifu zilibuka baada ya kupitia hatua bainifu (Hurskainen 2009a , 2009b, 2008, 2012, 2013a, 2013b 2014a, 2014b, 2014e, 2014f) katika utafsiri na utafiti wa mbinu hizi katika muktadha wa tungo-funge, vivumishi, vitenzi mbalimbali, virejehi, maswali, majina maalum na kadhalika. Pia tafsiri ya Biblia kutoka Kiswahili kwenda Kiganda Hurskainen (2009a) lugha yenye hazina haba za kikongoo. Michakato hii yote ya tafsiri-mashine imefafanuliwa zaidi kwa mifano na maelezo katika sura ya nne.

2.8 Programu ya Uwakifishaji ya - SWASENT

Hii ni programu inayotenda kazi mahususi za kuandaa matini kabla haziyapitia uchanganuzi wa programu tulizotanguliza kufafanua. Majukumu ya SWASENT ni kuwakifisha Kiswahili kwa kutia alama za uwakifishi, na alama za ziada za kufafanua maneno kilafudhi na kadhalika, kubadilisha herufi kubwa kuwa ndogo na kuhakikisha maneno ambayo yanafaa kuwa na herufi kubwa yamebakia hivyo, kuunganisha baadhi ya maneno yaliyoandikwa kando kando ila ni neno moja kisarufi kuwa neno moja k.m *wafanya kazi* kuwa *wafanyakazi* na kupanga matini vizuri ili kila sentensi moja iwe katika mstari wake na ianze katika nafasi yake na kadhalika (Hurskainen, 1999).

2.9 WLIST- Kuorodheshaji cha maneno

Ni ufupi wa Word-List kwa Kiingereza kumaana kuorodhesha maneno. Programu hii ya SALAMA hupanga matini za Kiswahili kwa mwegemeo wima kila neno katika mstari wake. Hurskainen (1999:141) anasema kuwa kuna aina nyingi za mbinu hii ambazo ni mahsusi kwa kila jukumu. Hakuna mengi yamesemwa kuhusu programu ya WLIST katika mbinu za SALAMA ila, inaonekana inaweza kusaidia mifumo mingi ya kuainisha, Kupunguza utata, kuainisha kisintaksia na kusaidiana

vizuri na programu za kompyuta kama **tokenisers** au vitenganisha-maneno kama alivyoshughulikia (Ng'ang'a- 2005) kama ilivyojadiliwa hapo juu.

2.10 Programu ya Kuunda na kutathmini Kamusi ya SALAMA-DC

SALAMA-DC inamaanisha *SALAMA-Dictionary Compiler* yaani kiunda kamusi cha SALAMA. Hili ni zao au matokeo ya ziada ya SALAMA na tofauti na programu na mbinu nyingine za kikompyuta za SALAMA zinazojisimamia zenyewe kama SWATWOL, SWACGP, na kadhalika tulizotangulia kuzifafanua hapo juu, *SALAMA-DC* huwa ni mfumo unaotumia programu hizo kadhaa kuunda na hata kutathmini kamusi ya Kiswahili.

Kwa mujibu wa Hurskainen 2008:3 SALAMA-DC ina vipashio vyote vinavyohitajika katika kutekeleza uchanganuzi au uainishaji wa lugha, inaweza kupokea matini-kavu (matini ambazo hazijaainishwa) kutoka katika sajili au nyanja yoyote ya lugha au ikiwa ni ya sajili au nyanja mahsusi kama vile sayansi, afya, elimu, siasa, kompyuta, kilimo kama (Sewangi 2001) na kwa hivyo ni programu inayofaa sana kwa uundaji wa kamusi ya Kiswahili kwa kuwa haijifungi na aina yoyote ya matini.

Katika kuunda kamusi lazima mtaalamu anayeshughulika aweze kuzalisha ¹⁵**vidahizo** yaani maneno makuu ya lugha husika na kuvielezea kikamilifu. Kila kamusi hulenga kuwakilisha msamiati wa lugha husika kwa ukamilifu iwezekanavyo. Hurskainen 2008, anaona kuwa hasa changamoto kuu si huku kuzalisha vidahizo vilivyo na habari tosha ya kiisimu bali zaidi ni majukumu magumu kama kuzaliwa mfumo wa vidahizo kurejeleana humo kwa humo na mfumo wa kioto wa kusimamia matumizi ya mifano wakilishi katika muktadha. Anatoa mahitaji kadhaa yanayotakiwa kutekelezwa ili shughuli hiyo ya mwisho –yaani mfumo wa kioto-ifanikiwe kikamilifu;

“...(a) Kwa lugha zilizo na muundo wa uambishaji mkubwa kama Kiswahili, mifano ni sharti itambuliwe kwa misingi ya muundo wa shina-lenye mzizi na viambishi ila si kwa muundo wa neno lilivyo katika matini. (b) ni sharti iwezekane kubainisha idadi ya juu zaidi ya mifano katika kila neno kuu au kidahizo. (c) Inafaa iwezekane kuhakikisha kuwa mifano inatolewa kutoka kwa kila aina ya matini asili. (d) Kwa kuwa katika kongoo-matini kuna idadi kubwa sana ya mifano, ni lazima kuhakikishwe kuwa idadi ya mifano inayotolewa inawakilisha uanuwai huo kikamilifu. (e) Tahadhari ichukuliwe ili hata maneno yasiyo ya kawaida yapate mifano ya kimatumizi katika kongoo. (f) Tungo-funge (kama nahau na methali) lazima zipate mifano ya kimatumizi. (g) Iwapo kidahizo au neno kuu ni homonimu (kama mbuzi – yaani leksimu mbili au zaidi katika lugha moja zenye kuandikwa sawa lakini zenye maana tofauti-(TUKI 1990:30), basi lazima kila kidahizo kipate mifano ya kimatumizi iwapo inapatikana katika kongoo-matini na kuwa lazima zifuatizwe homonimi husika katika nafasi yake baada tu ya kidahizo. (hili ni kuzuia utata wa kisemantiki kuzuka au kuepuka kuwachanganya watumiao kamusi hizo. (h) Pia ilazima iwezekane kufasiri urefu wa muktadha husika katika njia zilizomotishwa kiisimu, na sit u kufafanua idadi ya nambari au maneno kabla au baada ya kidahizo...”
(Hurskainen 2008:3) *Tafsiri yetu.*

Hoja hizi zitafanuliwa kwa kina katika Sura ya Nne katika utathmini wa data ya uundaji kamusi kutumia programu za SALAMA.

Kwa mujibu wa TUKI (2004:158) kidahizo ni neno linaloingizwa katika kamusi kwa chapa iliyokolezwa. Hata hivyo, fasili hii haina budi kutazamwa vema kwani si lazima kidahizo kiingizwe katika kamusi kwa chapa iliyokozwa. Huo ni mtindo tu wa uandishi; kila mwandishi anaweza kuwa na mtindo wake, mathalani, mwandishi anaweza kuamua kutumia chapa ya mlalo badala ya chapa iliyokolezwa kama ilivyoelezwa hapo juu. (Hans 2013:122)

Hata hivyo ni muhimu kusema hapa kuwa hatua zote nane zilizoordhereshwa katika nukuu hilo hapo juu zilitekelezwa katika *SALAMA-DC*. La muhimu zaidi hapa kama anavyotukumbusha Hurskainen, 2008 ni kuwa, licha ya kuwa mchakato huu wa kuunda kamusi kuwa unatumia mbinu za kikompyuta, hautekelezi kila jukumu kikamilifu kiasi cha kutohitaji mchango au uhariri wa binadamu, tunafaa kutambua kuwa sehemu kubwa na muhimu ambayo zamani ilichukua miezi na hata miaka na kazi ngumu za mikono ya binadamu sasa inaweza kufanywa kwa njia za kiteknolojia (kioto) na bado hatutakosa nafasi ya binadamu kutupia jicho katika kupiga msasa kamusi iliyozaliwa na mchakato huu.

Katika sura ya Nne tumeelezea kwa kina na kutolea mifano jinsi programu ya *SALAMA-DC* ifanyavyo kazi ya kuunda kamusi na kutathmini kamusi zilizopo.

2.11 Hitimisho

Sura hii imeshughulikia, lau kwa ufupi, programu mbalimbali zinazounda mradi wa *SALAMA* ambayo ni mazingira ya kughulikia lugha ya Kiswahili kwa kutumia mbinu za kikompyuta. Hapa tumedhihirisha kwa hivyo kuwa mradi wa *SALAMA* uliweza kuzamia utafiti na uchanganuzi wa lugha ya Kiswahili kwa mikabala anuwai inayompata mtaalamu wa lugha hii nafasi ya kushuhudia matumizi ya teknolojia ya lugha katika utafiti wake. Licha ya kuwa ujenzi wa programu za kompyuta ni mchakato wenye uchanganuzi wa hali ya kanuni za kialgorithmu, mtafiti ameweza kuangazia vipengele muhimu katika mradi huu kwa kutoa ufafanuzi unaoafiki malengo ya utafiti huu na kuongozwa na mipaka na mawanda ya utafiti wetu unaotumia *SALAMA* kama kielelezo cha matumizi ya teknolojia ya lugha katika Kiswahili.

Baadhi ya majukumu yameelezwa kuwa programu za *SALAMA* zinaweza kutekeleza ni pamoja na kuainisha lugha ya Kiswahili kimofolojia, kisintaksia na hata kisemantiki kwa kutambua na kueleza miundo ya ndani na ya nje kutumia kanuni za lugha zilizoandikwa kialgorithmu,, kusahihisha makosa ya tahajia za matini za Kiswahili, kutatua au angalau kupunguza utata wa maana ya maneno na pia wa kimofolojia, kuwakifisha matini za Kiswahili zilizo katika hali ya kompyuta, kutafsiri matini za Kiswahili hadi kwa lugha nyingine na kuhifadhi lugha ya Kiswahili. Kiambatisho 2, ukurasa wa 174 kinaonyesha kielelezo cha *SALAMA* na hatua mbali zinazopitiwa na matini katika utekelezaji wa baadhi ya mchakato iliyojadiliwa katika sura hii.

Hatimaye tumedhihirisha kuwa matumizi ya msamiati wa kutosha katika SALAMA umesaidia kuichambua na kuiainisha lugha ya Kiswahili kwa njia ambayo inaweza kusemekana ni kamilifu kwa kiasi kikubwa. Baadhi ya mapungufu yaliyojidhihirisha katika mradi wa SALAMA kwa mfano udhaifu wa kanuni kadhaa za kitatua utata wa kisemantiki na kadhalika, yalimulikwa katika sura ya nne kwa misingi ya nadharia zetu, maswali ya utafiti na haipotheisa za utafiti huu.

SURA YA TATU

3.0 KONGOO-MATINI KATIKA TEKNOLOJIA YA LUGHA

3.1 Utangulizi

Karne ya 21, imeshuhudia ukuaji mkubwa wa taaluma ya Isimu-Kongoo ambao ni mkabala wa kiisimu wa kushughulikia lugha kwa misingi ya kopasi au data ya lugha na pia ni tawi la Isimu-Kompyuta. Utafiti huu umeshirikisha katika malengo yake, haja ya kutathmini mchango wa kongoo na kuangalia jinsi kipengele hiki kinavyosaidia katika utafiti wa teknolojia ya lugha ya Kiswahili katika Mradi wa *SALAMA* na hatimaye inavyoweza kuwafaa watafiti wengine wa lugha hii. Pia tumetumia Mtazamo wa Kiisimu Kongoo kama mojawapo ya nadharia zetu. Katika sura iliyotangulia, tumetalii jinsi mradi wa *SALAMA* ulivyojenga mazingira faafu ya kikompyuta katika kushughulikia data hiyo ya Kiswahili kutumia programu za *kialgorithimu*. Lililobainika wazi katika Sura Ya Pili ni kuwa, ili programu yoyote iweze katambua sifa yoyote ya kiisimu ya lugha fulani, ni lazima pawe na hazina nzito ya msamiati wa lugha hiyo ambayo inawakilisha lugha katika hali yake halisi. Sura hii itatoa ufafanuzi wa kipengele hiki cha teknolojia ya lugha kinachoitwa kongoo na kufafanua jinsi kimetumika katika utafiti wa lugha ya Kiswahili.

Kanzi hii ndiyo tumeita kongoo au kopasi. Hurskainen (2005) anasisitiza kuwepo kwa kongoo kubwa iwezekanavyo na inayoshirikisha aina mbalimbali za maneno ili watafiti watumiao kompyuta waweze kuunda kanuni (mbinu) zifaazo za kuchanganua lugha. Sura hii imezamia kipengele hiki ikiongozwa na lengo lilotajwa hapo awali la kubainisha jinsi mradi wa *SALAMA* ulivyotumia amana hii inayojulikana kama Kongoo ya Kiswahili ya Helsinki (HCS-Helsinki Corpus of Swahili) katika kutekeleza utafiti wa teknolojia ya lugha ya Kiswahili. Hatimaye tumeangazia jinsi kongoo inavyoweza kutumika katika utafiti wa Kiswahili kwa jumla katika mazingira ya kikompyuta.

3.2 Maana ya Kongoo/Kopasi

Kwa mujibu wa Sinclair (2004), ambaye anafahamika kama baba wa Isimu Kongoo kutokana na utafiti wake katika eneo hili, neno katika lugha halina maana yoyote likiwa pekee yake ila maana hutokea pale maneno mengi huwa pamoja kwa mpangilio fulani. Neno ‘corpus’ kwa Kiingereza linatokana na neno la Kilatini ‘*corpus*’ linalomaanisha ‘*mwili*’. Hii ndiyo tunayoita kopasi au kongoo na hasa huwa mkusanyiko mkubwa wa matini ambayo ni kiwakilishi cha lugha fulani au sehemu yake ambao hutumiwa kwa uchambuzi wa lugha hiyo. (Sinclair 2004, Dash, 2010:35, McEnery and Wilson 1996:218). Wataalamu hawa na wengine wengi wanaojihusisha na uchunguzi wa lugha kwa mikabala ya kikongoo pia wanakubaliana na kipengele fulani katika fasiri ya maana ya kongoo, akiwemo Sewangi, 2001 kuwa kongoo ni mkusanyiko wa matini ambazo zinasomeka kikompyuta na huwa ni sampuli wakilishi za lugha halisi kwa kiwango kama itumiwavyo na

wahusika wake.

Naye Hurskainen 2005:122 anasema kuwa kongoo ni maandishi ya aina nyingi yaliyoandikwa siku baada ya siku na kukusanywa mwaka baada ya mwaka na kuhifadhiwa kikompyuta na huwa yanawakilisha lugha na matumizi yake kwa wakati uliohusika. Ni muhimu pia tuelewe pia kuwa kuna kongoo zilizoundwa kwa mkusanyiko wa maneno, si kama matini lakini kama sauti. Tutaeleza sehemu hiyo zaidi katika kijisehemu kimoja katika sura hii. Aidha mikusanyiko hii inaweza kuwa chimbuko lake ni magazeti, vitabu vya viada, vitabu mbalimbali, kutoka kwenye tovuti, vyombo vya habari na hata hotuba na mazungumzo halisi ya viongozi na wanajamii wengine (Hurskainen, 2005). Kwa ujumla wake, fasili ya kongoo kwa mujibu wa TUKI (1990:18), kuwa ni ‘...kusanyo la data za semi zilizoandikwa au kunukuliwa ambazo hutumiwa kama kianzio cha ufafanuzi wa kiisimu’...inapata mashiko na kutoa wazi sehemu muhimu za maana ya kongoo kama zilivyofasiliwa na wataalamu mbalimbali hapo juu.

3.3 Ukubwa wa Kongoo

Wataalamu wa Isimu-Kongoo kote ulimwenguni bado hawajawa na vigezo vya kuweka idadi kamili ya maneno yanayohitajika katika kongoo fulani ili iitwe kongo, ila wanakubaliana katika jambo moja; jinsi kongoo-matini ilivyo kubwa ndivyo inavyowakilisha lugha husika kwa njia ya kutegemewa zaidi. Kongoo kubwa huweza kushirikisha vipengele mbalimbali vya lugha husika na kuwakilisha nyanja, taaluma na sajili mbalimbali za lugha hiyo. Maendeleo ya teknolojia ya kompyuta yamechangia pakubwa katika dhana ya ukubwa wa kiongo kwa sababu ya uwezo wa (seva) kikompyuta wa kuhifadhi, kuchakata data na kuzifikia data hizo umepiga hatua kila uchao tangu miaka ya 1990 hadi karne hii ya ishirini na moja (Dash 2010). Miongo miwili iliyopita, ilikuwa inaaminika kuwa kongoo ya maneno milioni moja ilitosha kabisa kuwakilisha lugha fulani au sehemu fulani ya lugha hiyo.

Mambo sasa yamebadilishwa kabisa na kompyuta. Sasa kuna kongoo kubwa sana hasa za lugha za Uropa kama Kiingereza kwa kuwa lugha hizi zilianza kufanyiwa utafiti wa kikompyuta kwa muda mrefu kabla ya lugha nyinginezo. Kwa mfano, kongoo ya Kiingereza yenye maneno mengi zaidi ni kopasi ya ¹⁶COCA (*The Corpus of Contemporary American English*) yaani Kongoo ya Kingereza cha Kisasa cha Kimarekani ambayo ilikuwa na maneno milioni 425 kufikia 2011 na inaaminika kuwa inaweza kuwa hata imepita nusu bilioni (milioni 500) kufikia mwaka huu.

.....
Kongoo ya ¹⁶COCA iliundwa na mkusanyiko wa maneno ya Kiingereza cha Kimarekani kutoka mwaka 1990-2011 katika matini na sauti na huwa inapatikana kwa umma bila malipo katika <http://www.americancorpus.org>.

Pia kuna kongoo ya (*BNC-British National Corpus*) ambayo ni Kongoo ya Kitaifa ya Kiingereza cha Uingereza ambayo ilikuwa na maneno milioni 100 kufikia 2011 yakiwa ni asilimia 90 kongoo-matini na asilimia 10 kongoo-usemi (Nesselhauf, 2011) Pia huenda ikawa sasa imekuwa mara dufu au angalau kuongezeka zaidi miaka mine baadaye.

Lugha ya Kiswahili na lugha nyingine za Kiafrika ambazo kwa muda mrefu zimewekwa katika kategoria ya lugha ambazo wanaisimu-kongoo wanaita '*Lugha zenye hazina haba za kiutafiti*'. Lugha nyingi za Kiafrika zinaanguka katika kundi hili la 'jangwa la matini za kikompyuta' na Kiswahili kwa mujibu wa Kivaisi na Mbogho, 2012 ni mojawapo wa lugha hizi kwa kuwa wanadai kuwa bado hakina kongoo ifikishayo maneno milioni 100. Tunahisi dai hili sasa kuwa halina mashiko tena kwa kugemea maelezo ya Dash, 2010 hapo juu kuhusu uwakilishaji wa lugha wazo linaloungwa mkono na De Pauw na de Schryver (2009b). Kando na kuwa na hazina katika maandishi ya vitabu, mtandaoni na kwenye hifadhi za sauti, Kiswahili kinaweza kukwepa laana hiyo ya kupungukiwa na hazina za kidijitali.

Kwa sasa kongoo ya kutegemewa zaidi ya Kiswahili inayopatikana kwa utafiti wa teknolojia ya lugha ni Kongoo ya Kiswahili ya Helsinki (*HCS-Helsinki Corpus of Swahili*) iliyo na maneno ya Kiswahili sanifu na yaliyoainishwa kutumia programu za kikompyuta za SALAMA kwa mujibu wa Hurskainen, 2005 ilikuwa na maneno milioni 12.5 na inaendelea kuongezewa matini kwa sasa kukiwa na matini nyingi tu inayosubiri kuongezewa (¹⁷Hurskainen 2015). Kongoo hii ya Kiswahili ya Helsinki (HCS) itashughulikiwa kikamilifu baadaye kama sehemu na kipengele muhimu cha **SALAMA** katika sura hii na sura ifuatayo. Kongoo nyinginezo za lugha ya Kiswahili ambazo zinapatikana mtandaoni ni kama Kongoo ya TshwaneDJe Kiswahili Internet Corpus, TeDJe-KIC (De Schryver na Joffe 2009) ambayo ina maneno zaidi ya milioni 20 ila haijaainishwa kiisimu kama HCS. Pia kuna sehemu ya Kiswahili ya Kongoo Sambamba ya SAWA (SAWA Parallel Corpus) (De Pauw na wenzake. 2009), iliyokuwa na karibu maneno nusu milioni na pia kurasa 12, 000 za tovuti ya Wikipedia kwa Swahili zenye maneno zaidi ya milioni moja (De Pauw, G., de Schryver, G.-M. & Wagacha, Peter Waiganjo. (2009a).

De Pauw na de Schryver (2010:81), wakieleza kuhusu kongoo matini zinanavyokusanywa na umuhimu wa mikabala ya kikongoo katika utafiti, wanasema kuwa lugha zilizo na umuhimu wa kiuchumi ulimwenguni zimekuwa zikitumia mikabala hiyo kwa kuwa hazina hiyo ilipatikana, huku

¹⁷Hurskainen 2015- Nukuu zote za mwaka huu zitakuwa ni kutokana na majibu ya hojaji za utafiti hii na mawasiliano na mtaalamu huyu mwanzilishi wa SALAMA kwa barua pepe au njia nyinginezo katika mwaka huu wa utafiti.

lugha zilizo na hazina haba, hasa za kidijitali, zikibaki nyuma.

Hili limebakia kwa muda na kuwabidi wanaotafitia lugha hizi kubaki kutumia njia zisizo za kiyakinifu (kisayansi) zenye misingi ya kanuni na hivyo kujitenga na mikondo ya kisasa ya utafiti wa kiteknolojia ya lugha. Wanasema kuwa, lugha ya Kiswahili, licha ya kuwa inazungumzwa na mamilioni ya watu, bado haijaimarika vya kutosha katika nyanja ya kutafitiwa kiteknolojia.

3.4 Aina Za Kongoo

Kongoo za lugha ulimwenguni zimegawika katika makundi na wataalamu wa taaluma hii kutumia vigezo mbalimbali na wametofautia kwa mengi. Miongoni mwa wataalamu wa teknolojia ya lugha katika karne ya 21 waliojitwika jukumu la kuainisha kongoo za lugha kwenye makundi yake ni mwanaisimu-kongoo kutoka India Niladri Sekhar Dash ambaye aliweka kongoo kwenye makundi matano mapana ambayo ni kulingana na (i) Utanzu wa matini, ii) Aina ya data katika kongoo, iii) Aina ya Matini, iv) Lengo la Mkao wa Kongoo na (v) Namna ya Matumizi yake (Dash, 010). Tumetaja kwa ufupi yaliyomo katika kila kikundi cha kongoo na mifano yake kwa mujibu wa Dash.

3. 4.1. Utanzu wa matini

Kwa mujibu wa Dash, 2010 katika kitengo hiki tunapata aina tatu ndogo za kongoo ambazo ni (i) Kongoo-Matini: Kongoo iliyoandikwa kwa maneno (kwa mfano HCS- Kongoo ya Kiswahili ya Helsinki, Kongoo sambamba ya SAWA au Kongoo ya India ya TDIL Corpus) kwa kuwa utanzu huu unahusu tu data ya lugha iliyokusanywa kutoka kwa matini zilizoandikwa, zilizochapishwa, zilizokwenye majarida au vitabu au zilizotoka mtandaoni.

ii) Kongoo za Usemi zinashirikisha sauti za mazungumzo rasmi na yasiyorasmi, mijadala, mawasilisho, uchambuzi wa ghafla, mazungumzo ya kawaida, dialogia, monologia, hotuba kwa umma ambazo hazikupangiwa na kadhalika. Hapa vifaa vya kuwasilisha sauti hiyo haviangaliwi. Mfano wa Kongoo kama hii ni *Wellington Corpus of Spoken New Zealand English*) ya Kiingereza. (Dash, 2010:4).

iii) *Kongoo za Hotuba* Huwa na matini za kitaaluma zaidi ya Kongoo za Usemi kwa kuwa hotuba huwasilishwa katika muundo wa matini jinsi zilivyotolewa bila mabadiliko yoyote na kuchambuliwa kutumia aina ya programu ya kubadilisha matini hadi sauti. (Mfano wa kongoo ya aina hii ni *London-Lund Corpus of Spoken English*) ya Kiingereza kwa mujibu wa (Dash, 2010:4).

3. 4.2. Aina ya Data katika kongoo

Kundi hili linashirikisha vijishehemu vidogo sita katika kitengo hiki cha kongoo. Data huweza kugawika katika vijishehemu vifuatavyo;

(i) *Kongoo za jumla* ambazo huwa na matini ya jumla kutoka taaluma mbalimbali, tanzu, masomo na sajili tofauti. Aina hii ya kongoo inaweza kukua jinsi muda unavyosonga kwa kuongezewa

matini mpya. Huwa ni kubwa, ina uanuwai na uwakilishaji lugha mpana na matumizi mapana pia. (Mfano ni kama BNC- *British National Corpus*) ya Kiingereza tuliyoitaja.

(ii) *Kongoo spesheli*- imeundwa na sampuli mahsusi kutoka kwa kongoo za jumla ikiegemea aina fulani ya lugha hiyo kama lahaja, sajili, lugha ya watoto, lugha ya wasio wenyeji wake, lugha ya madaktari na kadhalika. Inafaa kwa utafiti kwa sababu sifa mahsusi ya lugha anayoitafuta mtafiti hujitokeza mara nyingi kuliko kongoo ya jumla (Dash, 2010:4). Kongoo kama hii huwa ni ndogo kuliko kongoo zinazoshirikisha nyanja zote za lugha. (Sinclair 1996b).

(iii) *Kongoo ya sehemu ya lugha* huhusisha matini moja ya aina moja tu ya lugha na hivyo huwa na data chache kuruhusu msamiati kujibana zaidi ya hata kongoo spesheli.

(iv) *Kongoo sampuli*- ni moja ya vijisehemu vya kongoo spesheli iliyoteuliwa kwa uangalifu kutoka kwa kijisehemu kimoja kwa minanjili ya kuchunguzwa. Unapoteua kongoo sampuli, haikuzwi kwa kuongezea maneno au kubadilishwa kwa vyovyote. (Sinclair 1991:24). Kwa kuwa mabadiliko yotote yanaweza kuvuruga muundo na mahitaji ya utafiti. (Mfano wa kongoo sampuli ni *Kongoo ya Zurich Corpus of English Newspapers*) ya Kiingereza.

(v) *Kongoo za kifasihi* ni kitengo spesheli cha kongoo sampuli ambapo vigezo vingi huwa vinahusishwa katika kuziainisha kama vile mtunzi wake, utanzu (riwaya, tamthilia, ushairi, hadithi fupi, wasifu n.k) mwaka, kundi, maudhui, na kadhalika. Kwa mfano, unaweza kuwa na kongoo ya mashairi ya Mahmud Abdilatif Abdulkadir (Ustadh Mau) ya baada ya miaka ya 2000. Mfano mwingine ni Kongoo ya riwaya za Shaban Robert inayoundwa na Taasisi ya Taaluma za Kiswahili-Chuo Kikuu cha Dar es Salaam (¹²Sewangi 2015).

(vi) *Kongoo fuatilizizi* ni kongoo inayozidi kukua kwa kuongezewa matini (k.m., *Bank of English*) ili iakisi ukuaji na mabadiliko ya lugha husika. Kongoo hii, inayoweza kuwa ni matini au ni sauti, na hufuatiliwa mwaka baada ya mwaka (Sinclair 1991: 21) ili kupata maneno mapya, mabadiliko ya matumizi ya maneno, kufuatilia msambao wa utokeaji wa misamiati. Baada ya muda habari inayoingia hudhibitiwa kwa sababu ya kasi iletwayo na mbinu mpya (Dash, 2010:5).

.....
¹⁹Sewangi 2015- Nukuu zote za mwaka huu za mtaalamu huyu zitakuwa kwa asilimia kubwa zaidi ni kutokana na majibu ya hojaji za utafiti hii na mawasiliano naye kwa barua pepe au njia nyinginezo katika mwaka huu wa utafiti.

3.4.3. Aina ya Matini katika kongoo (Type of Text)

Kongoo za aina hii huwagika tena kwenda kutatu (Dash, 2010:5) nazo ni (i) Kongoo ya lugha moja - huwa na matini zinazowakilisha matumizi ya lugha moja katika kipindi cha muda fulani na inaweza kuwa matini na sauti bora hazihitilafiani. (Mfano wa kongoo kama hii ni *ISI Bengali Corpus*) inayowakilisha lugha ya Bengali nchini India.

(ii) Kongoo ya lugha mbili (Bilingual corpus) huwa imeundwa wakati kongoo za lugha mbili zinazohusiana au ambazo hazihusiani huwekwa kwa hifadhi moja. Iwapo lugha hizi zinahusiana kiotografia/kimaandishi au kinasaba basi kongoo hizi zinaweza kuitwa Kongoo Sambamba (Mifano ya kongoo hizi ni kama (Kongoo ya *TDIL Bengali-Oriya Corpus* ambayo ina lugha mbili za India za Bengali na Oriya (Dash, 2010:5) na pia kama tulivyodokezea kuna kongoo sambamba ya SAWA ya Kiswahili na Kingereza (²⁰SAWA Corpus) iliyo na idadi ya maneno milioni moja yaliyokusanywa kutoka kwa nyaraka za kisiasa, matini za kidini, maandishi ya filamu, ripoti za uwekezaji na nyaraka nyingine katika Kiingereza na Kiswahili. (De Pauw na wenzake. 2009).

(iii) Kongoo ya Lugha Nyingi - Kongoo za iana hii huwa zina mkusanyiko wa matini wakilishi kutoka kwa lugha zaidi ya mbili na kama kongoo sambamba huwa kategoria za matini na mbinu za kupata sampuli ya matini za lugha husika zinasawazishwa (Dash, 2010:5). (Kongoo za iana hiyo ni kama *Oslo Multilingual Corpus-OMC* –iliyo na lugha za Kiingereza, Kinorweje, Kifini, Kijerumani na Kireno, Kiswedi na Kiholanzi na kuna mipango ya kuongezea Kibelgiji na Kifaransa) Johanson, Stig (2010) na pia kuna (*Crater Corpus*).

3.4.4 Lengo la Mkao wa Kongoo (Purpose of Design)

Kundi hili linahusisha aina mbili za kongoo ambazo ni (i) Kongoo kavu/isiyoalamishwa (Un-annotated corpus): inayowakilisha matini rahisi katika hali yake asilia bila kuongeza habari zaidi za uchanganuzi wa kiisimu au usio wa kiisimu. Inafaa kwa utafiti wa lugha ila kongoo-matini ambayo imeainishwa inaongeza thamani katika utafiti (Dash, 2010:5) na (ii) Kongoo iliyoainishwa/Iliyoalamishwa (Annotated corpus): Hii ni kongoo ambayo inao ufafanuzi na

²⁰SAWA Corpus- (De Pauw and Wagacha, 2007; De Pauw et al., 2009:11-12) ni kongoo sambamba au kongoo pacha ya Kingereza na Kiswahili ambayo ilitumiwa katika utafiti wa tafsiri-mashine na wataalamu hawa watatu utafiti wa tafsiri-mashine. Kongoo hii ilijumuisha maneno zaidi ya nusu milioni ya Kiingereza na karibu idadi hiyo kwa Kiswahili yalitolewa kwa lugha hizo mbili kutoka kwa vianzo vifuatavyo; Biblia-Agano jipya-151,100, Quran- 124,300, Maafikiano ya Haki za Kibinadamu-1,800, tovuti ya Kamusi.org-26,700, Maneno ya tafsiri kutoka kwa filamu teule- 58,400, Ripoti za uwekezaji- 54,900 na kutoka kwa watafsiri wa kibinafsi kama Dkt Zaja Omboga na wengineo ni 25,700. Wanateknolojia Lugha hawa, walipendekeza kuwa, iwapo matini zilizotafsiriwa na zinazolingana katika lugha hizi mbili zinaweza kuongezewa na uainishaji wa kina zaidi kufanyika, basi mfumo wa walioujaribu wa tafsiri-mashine, utafaa sana kwa kazi za Kiswahili kwenda Kiingereza au Kiingereza kwenda Kiswahili. Kongoo sambamba husaidia sana katika taaluma ya tafsiri-mashine kwa kutumia teknolojia ya lugha na hasa TM yaani hifadhi ya maneno yaliyotafsiriwa ya lugha ya pili.

ishara zilizopachikwa kutoka nje na waundaji wake ili kuleta habari zaidi (kama alama za uchanganuzi, alama za kuonesha aina za maneno, habari za kategoria za kisarufi) kwenye matini. Kongoo ya aina hii, tofauti na kongoo kavu, huwa inafaa zaidi katika kutoa habari muhimu kwa majukumu tofauti katika teknolojia ya lugha pamoja na uchakataji wa kimofolojia, uchambuzi wa kikompyuta wa sentensi, kupata habari muhimu kutoka kwa kongoo, kupunguza au kuondoa utata wa maana za maneno, tafsiri-mashine na kadhalika (Dash, 2010:5). (Mfano wa Kongoo kama hii ni British National Corpus ya Kiingereza na HCS-*Helsinki Corpus of Swahili* yaani Kongoo ya Kiswahili ya Helsinki ambayo tumeimakinikia zaidi baadaye katika sura ya nne.

Uandaaji au uchakataji wa matini katika kongoo ni mchakato muhimu sana katika utafiti wa lugha kutumia teknolojia ya lugha ili kuzalisha data bainifu ya lugha husika. Pindi kongoo inapoundwa huwa kuna haja ya kuibuka na mifumo, zana za kikompyuta, mbinu na programu za kutathmini data ya lugha na kutoa habari muhimu kutoka kwa kongoo (Nash, 2010:12). Kunazo mbinu kadhaa za uandaaji wa kongoo kama vile uchanganuzi wa kitakwimu(statistical analysis), upatanisho wa kisarufi katika matini (concordances), ujirani wa misamiati mbalimbali (lexical collocations), usakuzi wa maneno makuu (key word search), ukundishaji wa maneno ya kawaida (local-word-grouping), utengaji wa mzizi wa neno (lemmatization), uchambuzi wa kimofolojia (morphological processing), uzalishaji maneno(generation), utambuzi wa aina za maneno (*parts-of-speech-tagging*) ualamishaji (annotation), uchanganuzi (parsing) na kadhalika ambazo zimedhihirisha kuwa mbinu za kongoo huwa aghalabu zinakinzana na mbinu za utafiti za kimapokeo kama za kihisia-ng'amuzi kuhusu lugha na sifa zake (Dash,2010:12). Miktadha ya matumizi ya lugha ya Kiswahili kutoka kwa kongoo ya HCS iliyofafanuliwa na kutolewa mifano zaidi katika sura ya nne ni ushahidi tosha.

3.4.5 Aina ya Matumizi ya Kongoo (Nature of Application)

Kwa mujibu wa Dash 2010, katika kundi hili la kongoo kuna vikundi sita vidogo ambavyo tutavitaja kwa ufupi tu. (i) Kongoo za Tafsiri (Translation Corpora): Hizi huwa na matini kutoka lugha chanzi na tafsiri zake katika lugha pokezi. Kwa kuwa kongoo za aina hii hutoa msingi imara wa kulinganisha hali za kiisimu katika lugha husika katika mazingira yanayofanana, basi zinachukuliwa kama hazina faafu kwa uchanganuzi wa data baina ya lugha mbili na uundaji wa kanuni za kikompyuta zihitajikazo kwa taaluma ya tafsiri. (Altenberg and Aijmer 2000:17).

(ii) *Kongoo iliyooanishwa (Aligned corpus)* huwa ni aina ya kongoo ya lugha mbili ila huwa matini kutoka kwa lugha moja na tafsiri kwa lugha pokezi huwa zimeoanishwa kisentensi, kishazi kwa kishazi, na neno kwa neno. (Mfano wa kongoo ya aina hii ni (*The Canadian Hansard Corpus*) (Dash 2010:5) ya Kiingereza na Kifaransa. Nchi zilizo na sera za uwili-lugha zinafaa kuwa na

nyaraka rasmi katika mpangilio huu wa kongoo iliyoanishwa ili ziyakidhi masharti ya kisheria na kimawasiliano. (iii) Kongoo Sambamba (Parallel corpus): Aina hii ya kongoo inaweza kuchukua maelezo ya kongoo iliyoainishwa hapo juu ila hii pia inaruhusu kuhakikisha visawe vya tafsiri za lugha husika. Mfano wa kongoo hii ni ile ya *Chemnitz German-English Corpus*) ya Kijerumani na Kiingereza.

(iv) *Kongoo Marejeleo (Reference corpus)*: Hii ni kongoo iliyo na sifa kama za Kongoo ya Jumla tuliyoitaja hapo juu kwa kuwa ni kubwa, inawakilisha nyanja tofauti za lugha, na hutumiwa kuandikia sarufi, kamusi, thesauri, na machapisho mengine. Huwa ni kigezo cha msamiati wa lugha husika na huaminiwa na jamii lugha husika. Pia hutathmini msamiati unaoshughulikiwa na mbinu na programu za teknolojia ya lugha. Inajumuisha matini andishi, sauti, lugha rasmi, isiyo na hata sajili za jamii na za hali fulani. (Mfano ni kongoo ya *Bank of English* kwa Kiingereza). Aidha tunaweza kudai kuwa Kongoo ya HCS ya Kiswahili ina baadhi ya sifa hizi na karibu inafikia viwango vya kundi hili la kongoo.

(v) *Kongoo Linganishi (Comparable corpus)*: huu ni mkusanyiko wa matini ‘zinazofanana’ kutoka kwa lugha zaidi ya mbili. Ni matini kutoka kwa lugha tofauti zinazotumiwa kuzilinganisha lugha hizi. (Mfano mwafaka wa kongoo hii ni *Corpus of European Union*).

(vi) *Kongoo za Ki-nafasi/Ki-abra/Ki-fursa (Opportunistic corpus)* Ni kongoo za kielektroniki zinazokusanywa kwa njia isiyo ghali, hutumiwa bila malipo au kwa bei nafuu ila aghalabu huwa si kamilifu. Watumiaji hubakia kujaza mapengo wenyewe. Huwa kama kongoo fuatilizi na mtafiti hutumia kulingana na mahitaji ya utafiti wake (Dash, 2010:6).

3.5 Sifa kuu na muhimu za kongoo

Mwanaisimu-kongoo (Dash, 2010:2) alipendekeza sifa bainifu za kongoo ya lugha yoyote ambazo alisema kuwa zinafaa kuwa vigezo vya kupimia ukamilifu au ufaafu wa kongoo ya lugha. Tumeamua kufuata mkabala wa Dash kwa kuwa ni shirikishi kimaielezo na bado hatujapata mwana-isimu mwingine mwenye maoni kinzani kuhusu sifa za kongoo. Sehemu hii ni muhimu kwa utafiti huu kwa kuwa itasaidia katika vigezo na misingi ya kutuelekeza wakati tutakapotathmini kongoo ya HCS kama chombo cha teknolojia ya lugha na ufaafu wake katika utafiti wa Kiswahili kwenye Sura ya Nne.

3.5.1 Ukubwa

Inapaswa kubwa na ya kutosha uwakilishaji wa lugha husika kwa kusheheni data kwa muundo wa matini au sauti. Ukubwa hapa humaanisha ujumla wa idadi ya vipengele vyake vyote vinavyounda kopasi nzima na kuwakilisha lugha.

3.5.2. Ubora wa kongoo (uasilia)

Matini zote ziwe zimetokana na mifano halisi ya mazungumzo au maandishi. Nafasi ya mwanaisimu hapa ni muhimu kwa kuwa lazima athibitisha kuwa data hii ya lugha imekusanywa kutoka kwa mawasiliano ya kawaida wala si kutoka kwa hali za kimajaribio au katika hali za kuundwa au kuigizwa na binadamu. **(Dash, 2010:2).**

3.5.3 Uwakilishaji

Inapaswa kuwa inashirikisha sampuli mbalimbali kutoka kwa matini za aina tofauti. Inapaswa kuwa imewiana na sehemu zote za matumizi ya lugha ili iweze kuwakilisha kwa kiasi kikubwa iwezekanavyo uanuwai wa kiisimu katika lugha hiyo, kwa kuwa uchambuzi wa baadaye utakaojikita katika kongoo hii, unahitaji kuthibitisha na kuhakikisha habari za kila aina katika kongoo ambayo inawakilisha lugha. **(Dash, 2010:2).**

3.5.4 Usawa

Sampuli zilizotumika katika kongoo ziwe za viwango sawa. Hata hivyo, hili ni suala lenye utata kwa kuwa haliwezi kutekelezwa kila mahali. Mbinu ya ukusanyaji sampuli zinaweza kubadilika kwa kiasi kikubwa ili kuifanya kongoo iwe na uwakilishaji wa kutosha au kuwa inayofumbata nyanja kadhaa.

3.5.5 Urahisi

Kongoo inapaswa kuwa na matini tupu katika muundo ulio rahisi kabisa. Hii inamaanisha kuwa inatarajiwa kuwa na mfuatano wa herufi au maneno bila viziada vingine vya kiisimu kuongezewa katika matini. Kuwa na matini ya kimsingi ni tofauti na kusema kuwa kuna aina ya uainishaji wowote ulio na habari za kiisimu au zisizo za kiisimu (Dash, 2010:2). Mfumo wa kuweka lugha katika matini kavu (native text) ni mfumo wa Linux kwa kuwa mifumo mingine ina viziada tusivyoona kama vile Ms word (Hurskainen, 2008). Pia wazo hili la usahili wa matini za kongoo umeungwa mkono na (Sinclair, 2004).

3. 5.6 Upatikanaji au Urahisi wa kupatikana

Watumiaji wa kongoo wanapaswa kuwa na urahisi wa kupatikana tena kwa data, habari, mifano na marejeleo kutoka kwa kongoo na hili huwezesha na mbinu mwafaka za kuhifadhi data ya lugha katika mfumo wa kielekroniki kwa kuwa uwezo wa kielekroniki unawezesha haya katika kompyuta kwa sasa. (Dash, 2010:2).

3.5.7 Uthibitishikaji

Kongoo inapaswa kuwa wazi kwa uthibitisho wa kiuyakinifu (empirical verification). Tunaweza kutumia data kutoka kwa kongoo kwa uthibitisho wa aina yoyote ya hali ya lugha. Sifa hii huweka Isimu-Kongoo kidedea na hatua nyingi mbele ya mtazamo wa kiung'amuzi (intuitive approach) katika kuchunguza au kufanyia utafiti lugha.

3.5.8 Mpangilio wa habari kwenye kongoo (Mpangilio wa habari kwenye kongoo)

Habari kamili ya vijenzi au vijishehemu vya kongoo, vilivyo na mada ndogo za kurejelea uhifadhi nyaraka, inafaa kutengwa mbali na matini yenyewe ili kusaidia utenganishaji wa haraka wa matini na viainishi vyake bila kutumia juhudi za programu nzito za kikompyuta. (Dash, 2010:2).

3.5.9 Uongezekaji

Kongoo inapaswa kuwa inaongezeka mara kwa mara jinsi muda unavyosonga. Sifa hii huiweka kongoo katika hali sambamba na mabadiliko ya sajili za lugha yanayotokea mara kwa mara. Kwa kuongezeka kwa data ya lugha, kongoo huweza kufaa kwa utafiti wa kihistoria na mapisi ili kukumbana na mabadiliko katika jamii na katika maisha (Dash, 2010:2).

3.6 Kongoo Ya Kiswahili ya Helsinki- SALAMA (HCS)

Kongoo ya Kiswahili ya Helsinki almaarufu kama Helsinki Corpus of Swahili (HCS) ni kongoo matini ya Kiswahili sanifu iliyoainishwa kiisimu kutumia programu za SALAMA. Kongoo ya HCS ina takriban maneno milioni 12.5 na iliundwa kutokana na matini za magazeti ya Kiswahili kutoka Kenya na Tanzania kama vile Kasheshe, Majira, Nipashe, Uhuru, Komesha, Mfanyikazi, Mzalendo, Alasiri, Rai, Lengo, Al nuur na Majira na tovuti za habari za Deutsche Welle pamoja na vijishehemu vya matini zilizotolewa kutoka kwa vitabu na machapisho ya kinathari ikiwa ni pamoja na vitabu vya kibunilizi, vya kielimu na vya kisayansi kama itakavyoonyeshwa katika jedwali la Jedwali 3 (Hurskainen, 2004). Kwa mujibu wa Kivaisi na Mbogho (2010) hii ndiyo kongoo kubwa zaidi ya Kiswahili iliyoainishwa inayoweza kupatikana mtandaoni katika karne ya ishirini na moja. Kongoo ya HCS inapatikana kwa utafiti kwa kutuma maombi yako katika chuo Kikuu cha Helsinki kituo cha kompyuta kama tulivyosema hawa awali.

Kongoo ya HCS ilichanganuliwa na SALAMA (Swahili Language Manager), ambacho, kama tulivyoeleza katika sura iliyotangualia ni chombo chenye majukumu anuwai ya kushughulikia lugha kilichoasisiwa na Arvi Hurskainen, Profesa wa lugha za Kiafrika katika chuo kikuu cha Helsinki. Kongoo hii pia inahusisha matini kutoka kwa misahafu kama vile (Biblia na Quran) (Ng'anga 2005). Taarifa ya maneno ya Kiswahili inayopatikana katika kongoo hii ni kama mzizi wa neno (lemma) aina ya neno, uchanganuzi wa kimofolojia pamoja na uhusishwaji wa ki-ngeli za nomino

husika na mofolojia ya kitenzi husika, etimolojia ya maneno yaliyokopwa na pia faharasa yake kwa Kingereza.

3.6.1 Vyanzo vya Kongoo ya Helsinki Corpus of Swahili (HCS)

Kongoo hii ina matini za Kiswahili sanifu kutoka kwa vyanzo mbalimbali kama ilivyodokezewa hapo awali;

- i) Matini tofauti kutoka kwa magazeti mbalimbali ya Kiswahili nchini Tanzania kama vile *Mzalendo, Uhuru, Nipashe, Majira, Rai, Kiongozi, Alasiri, Al nuur, Lengo n.k.* Pia kuna matini kutoka kwa tovuti ya habari kwa Kiswahili kama vile *Deutsche Welle* – kutoka Ujerumani.
- ii) Makala zilizokusanywa kutoka kwa vitabu mbalimbali vya Kiswahili kama vile vya fasihi kama vitabu vya waandishi kama Said Ahmed Mohammed, Mohammed S. Mohammed, Euphrase Kezilahabi na vitabu vingine vya sarufi.
- iii) Mkusanyiko wa hotuba na mahojiano katika lahaja tofauti za Kiswahili kutoka visiwa vya Pemba, Zanzibar, Timbatu na Mafia kwa ushirikiano wa karibu wa TUKI na Chuo Kikuu cha Helsinki katika mradi wa DAHE.
- iv) Kongoo ya HCS pia inajumuisha misaafu yaani vitabu vya dini kama *Bibilia* ya Kiswahili na Quran kwa Kiswahili. Makundi kama *United Bible Society* yalichangia pakubwa katika kufaulu kwa upatikanaji wa matini hizi.
- v) HCS pia inajumuisha makamusi mbalimbali ya Kiswahili pamoja na *Kamusi ya Kiswahili Sanifu n.k.* miongoni mwa matini nyinginezo nyingi za Kiswahili.

H.C.S ni mradi unaozidi kupanuka kwani matini mpya bado zinazidi kuongezwa hasa kupitia kwa mtandao na vyanzo vinginevyo kama asemavyo (Hurskainen, 2015). Hurskainen, 2002:286 pia alipotumia kongoo hii katika utathmini wa Kamusi tano za Kiswahili, anathibitisha uwepo wa matini hizo alipogawa kongoo alizotumia kutathminia kamusi hizo kama ifuatavyo alivyoviita

²¹Korpus 1, Korpus 2 na Korpus 3. Matini hizo ni:

- i) Kongoo ya magazeti yanayopatikana katika Internet, pamoja na habari za Deutsche Welle kwa Kiswahili (Korpus 1). Data hizo ni kutoka miaka 1998-2001, na jumla ya maneno ni 2.484.852.
- ii) Kongoo ya magazeti ya zamani zaidi (miaka 1988-1994) pamoja na vitabu mbalimbali (jumla ya vitabu kama 45 hivi. Korpus 2).
- iii) Korpus hii inayo maneno 1,190,489 Kongoo ya magazeti ya kisasa (mwaka 2002) pamoja na vitabu kadhaa (Korpus 3), jumla ya maneno ni 552.02.

²¹Korpus 1, Korpus 2 na Korpus 3 ni vitambulisho vya kongoo tatu mahsus zilizofafanuliwa hapo juu pamoja na vyanzo vyake alivyotumia Hurskainen, 2002 katika utathmini wa kamusi tano tofauti za Kiswahili kutumia programu za SALAMA hasa SALAMA-DC.

Ithibati nyingine inayoonyesha kuwa kongoo ya HCS imeundwa na matini kutoka kwa vitabu vya kifasihi ni katika Hurskainen, 2008:15 ambapo asema kuwa alijaribu programu ya kiunda-kamusi cha SALAMA-DC kutumia sehemu ya kongoo iliyotokana na vitabu vitano vya mwandishi E. Kezilahabi iliyokuwa na maneno 196,150. Vitabu hivi ni baadhi ya matini za Helsinki Corpus of Swahili. Kamusi iliyopatikana katika mchakato huu ilikuwa na vidahizo vya maneno hayo yote yakiwa yamepangwa katika orodha ya mistari 18,750.

Hurskainen (2009), katika maelezo kwenye tovuti ya SALAMA ameonyesha muundo wa kongoo ya HCS ilivyokuwa kufikia mwaka huo kulingana na vyanzo vya matini na mgao kwenye tovuti pamoja na idadi ya maneno katika vitengo mbalimbali, kama ilivyo katika jedwali lifuatalo;

	Vyanzo vya Matini	Jina la Kitengo kwenye HCS	Idadi ya makala	Idadi ya Maneno
1	Alasiri	articles/alasiri/	2779	125958
2	Al nuur	articles/annuur/	659	837990
3	Vitabu	books/	72	105542
4	Tovuti ya Dwelle	articles/dwelle/	9831	2479606
5	Kasheshe	articles/kasheshe/	7	15388
6	Kiongozi	articles/kiongozi/	70	828256
7	Komesha	articles/komesha/	1	8948
8	Lengo	articles/lengo/	6	2, 347
9	Majira	articles/majira/	6992	3,309,197
10	Mfanyakazi	articles/mfanyakazi/	19	7,503
11	Mzalendo	articles/mzalendo/	451	366,307
12	Nipashe	articles/nipashe/	4,086	2,019,471
13	Rai	articles/rai/	33	242,281
14	Uhuru	articles/uhuru/	816	311,581

Jedwali 3: Migao ya matini za chimbuko la Kongoo ya Helsinki Corpus of Swahili

3.6.2 Historia ya Kongoo ya Kiswahili ya Helsinki (HCS)

Kongoo ya HCS ambayo ni kongoo matini ilianza kukusanywa katika mwisho wa miaka ya 1980 katika Chuo Kikuu cha Dar es Salaam na wanafunzi wajioshirikishwa katika mradi huo kwa minanjili ya utafiti wa SALAMA (Hurskainen, 1999). Ukusanyaji wa matini hizi uliitwa DAHE kama tulivyotaja mwanzoni ambao ni mradi wa ushirikiano baina ya Chuo Kikuu cha Dar es Salaam na Chuo Kikuu cha Helsinki. Kongoo ya HCS yenyewe kama mradi wa kiutafiti ni mali ya taasisi ya masomo ya Kiasia na Kiafrika katika chuo kikuu cha Helsinki – Finland ukiongozwa na Arvi Hurskainen, kama ilivyotajwa hapo awali.

Kando na matini zilizotajwa hapo juu za kutoka kwa magazeti na tovuti za Kiswahili, kongoo ya HCS pia imeundwa kutokana na sauti za mazungumzo yaliyokusanywa na kuwekwa katika hali ya matini kutoka mwaka 1989-1991 kutoka kwa wazungumzaji wa eneo la Pwani ya Tanzania na kutoka visiwa Zanzibar, Pemba na Mafia. Sehemu hii ya kongoo pia haijaalamishwa na wanaoomba kibali cha kuitumia kwa malengo ya kuitafiti, hupewa kupitia kwa jukwaa la Linux. Ombi hili hutumwa kando na la Kongoo matini ya HCS ya jumla ambayo tofauti na hii iliyoandaliwa kutoka kwa sauti na haijaalamishwa, hutumia jukwaa la Lemmie. (Hurskainen, 2007, 2009).

Kongoo ya HCS imekuwa uwanja wa utafiti wa masuala ya teknolojia ya lugha ya Kiswahili na wasomi mbalimbali wakiwemo wanasayansi ya kompyuta, wanaisimu, wataalamu wa Kiswahili na hata wataalamu wa tafsiri-mashine. Wataalamu waliotangamana na kongoo hii katika kazi zao za kiutafiti katika kiwango cha shahada ya uzamifu ni Dkt. Wanjiku Ng'ang'a (2005) kutoka Kenya na Dkt. Suleiman Sewangi (2001) kutoka Tanzania.

Kufikia mwaka wa 2004, Helsinki Corpus of Swahili ambayo ilipatikana katika seva ya Helsinki University Language Corpus, iliamishwa hadi kwa seva ya kongoo ya ¹⁶CSC yaani Scientific Center for Computing (CSC). Kongoo hii ambayo kitambo ilikuwa kongoo-kavu, yaani bila kualamishwa-sasa imealamishwa na iko katika muundo wa XML na inazo alama za kiisimu. Iliainishwa au kualamishwa kikompyuta kutumia programu za SALAMA na hali ya sasa huiwezesha kusaka na kupata habari tofauti tofauti kuhusu lugha (Hurskainen, 2007). ²²Mwongozo wa alama za kiisimu zilizotumiwa kualamisha kongoo ya HCS kwenye tovuti kama ilivyodokezwa katika sura ya pili.

.....
²²Mwongozo wa alama za SWATWOL zilizotumika katika kongoo ya SALAMA hupatikana katika <http://www.aakkl.helsinki.fi/cameel/corpus/swatags.pdf>

Tunaelewa kwa kongoo ya HCS bado inaendelea kurubishwa kwa kuongezewa hazina Zaidi za matini ya Kiswahili kwa mujibu wa Hurskainen 2015;

“...Nimekusanya matini mengi zaidi, lakini hatujafanikiwa kupata msaada wa kifedha kwa kuzishughulikia ili kuziingiza kwenye kongoo. Pia inatakiwa kongoo nzima iainishwe upya kwa kutumia **SWATWOL** ya kisasa...” (12**Hurskainen, 2015**).

Kwa hivyo ni dhahiri kuwa kazi ya teknolojia ya lugha katika mradi wa SALAMA ingali inaendelea na kuwa haja ya kuzidi kuongezea matini katika kongoo hii ili iweze kufumbata hali halisi za mabadiliko ya lugha na uibukaji wa istilahi mpya.

3.6.3 Masharti ya kuitumia kongoo ya Kiswahili ya Helsinki (HCS)

Kongoo ya Helsinki Corpus of Swahili (HCS) inaruhusiwa kutumiwa na watafiti katika vyuo na taasisi za elimu na utafiti mradi upate idhini ya kujisajili mtandaoni ili kuruhusiwa kuifikia matini katika kongoo hii kwa muda wa miaka miwili. Akaunti ya kutumia kongoo hii huwa ni ya mtu au taasisi binafsi na haipaswi kupeanwa kwa mhusika wa tatu. Masharti ya matumizi ya kongoo ya HCS yanafaa kuambatana na sheria za haki miliki za Ufini na za kimataifa kwa mujibu wa **Hurskainen (2009)**.

Wanaoomba kutumia matini za kongoo ya HCS wanafaa kufanya hivyo tu kwa malengo ya kiusomi na uchanganuzi wa kiisimu. Kongoo hii au sehemu yake yairuhusiwi kuchapishwa katika vitabu, majarida au kwa njia za kielekroniki au mtandaoni. Aidha, wanaoitumia katika utafiti wao wanafaa kuinukuu ipasavyo. Kwa mfano, iwapo mtafiti atumia kongoo ya HCS katika mwaka wa 2015 anafaa kunukuu; **HCS 2015. Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC–IT Center for Science. (Hurskainen, 2004)**.

Viwango vinavyokubalika vya unakili wa kazi za kitaaluma kwa matumizi ya kiutafiti au kiufundishaji vinafaa kutumika almradi wanaotumia nakala za kongoo hii hawazisambazi au kufanya zipatikane katika muundo wowote kwa mfano kwenye intaneti. Iwapo nakala za kongoo ya **HCS** alizotumia mtafiti zimetolewa chapa, basi zinafaa kuharibiwa baada ya kumtimizia haja zake za kiutafiti. Tahadhari pia imetolewa kwa watafiti kuwa licha ya kuwa matini katika **HCS** ilishughulikiwa kwa makini yote na masahihisho na marekebisho kufanywa, ualamishaji na uchanganuzi wa kongoo hii ulifanywa kwa njia za kikompyuta kwa hivyo kuna uwezekano wa kukumbana na mategu ya kitahajia hapa na pale.

Kibali cha kutumia kongoo hii kiutafiti hukamilika baada ya miaka miwili kutoka siku ya kapata akaunti ya kutumia au punde tu mradi wa utafiti unaohusika unapokamilika au wakati mhusika anapotamatisha muda wake katika chuo au taasisi ya masomo. Masharti mengine ya kimatumizi

katika kongoo matini ya HCS ni kwamba lazima kongoo itumike kwa jukumu la kiusomi lililopendekezwa awali, kupelekeana matini hizi kwa mitandaoni kuwe ni kwa idhini ya idara ya CSC na walioruhusiwa kutumia nje ya nchi ya Finland wanafaa kulinda akaunti zao kwa nywila zisizotabirika. Iwapo mtafiti anakaidi masharti haya, mkataba wake wa kutumia kongoo hii pia unafutiliwa mbali. (Hurskainen, 2004, 2007). Habari tangulizi zinapatikana katika wavuti wa kongoo hii katika <http://www.aakkl.helsinki.fi/comeel/corpus/intro.html>

3.7 Nafasi ya Kongoo ya HCS katika Mradi wa SALAMA

Kwa mujibu wa Hurskainen 2009, maagizo yaliyotolewa katika tovuti ya SALAMA ni ya jinsi ya kutumia kongoo ya HCS katika mifumo ya Unix/Linux na jinsi ya kutumia kongoo hiyo pamoja na programuau mbinu za kawaida.²¹Maagizo haya hupatikana kwenye tovuti ya SALAMA.

Kwa upande mwingine SALAMA, ni mazingira changamano sana kwa matumizi ya kongoo na huwezesha uendeshaji, ushughulikiaji bayana na usimamizi ufaao wa habari za lugha Kongoo ya HCS huwa inaalamishwa huku mtafiti akifanya kazi nayo. Hivi ni kusema kuwa unaweza kuainisha na matini za kongoo na kisha unatumia matokeo ya uchanganuzi huo wakati unafanya kazi na kongoo hiyo. Hii ni sawa na kufanya kazi na kongoo iliyoalamishwa. Vyombo vya uchanganuzi na mfumo wa alama za kuainisha kongoo katika SALAMA kutumia SWATWOL vinavyopatikana kwenye tovuti ya SALAMA.

Hurskainen 2009, anatoa sababu za kutoainisha kongoo yote ya HCS kuwa kongoo iliyoainishwa huchukua nafasi kubwa sana hata zaidi ya mara kumi kwenye seva ya kompyuta, mtafiti anaweza kuainisha kongoo hii kutumia zana za SALAMA anapohitaji kongoo ya aina hiyo, kuna aina mbalimbali za uainishaji kwa hivyo zote zikitumiwa zitatumia nafasi kubwa kwenye kompyuta na hivyo haja ya mtafiti kujaiinishia mwenye na mwisho ni kuwa kushughulikia lugha kwa kutumia kongoo isiyoainishwa kunampa mtafiti nafasi nzuri na mnepo wa kikazi katika utafiti wake.

Katika kuchanganua kongoo ya HCS kutumia programu za SALAMA, michakato yote ya kuichunguza lugha inazo taratibu zake. Michakato hii inayofanyiwa kongoo, ambayo hapa huchukuliwa ndiyo lugha yenyewe, inaweza kutumika katika uchanganuzi wa kimofolojia, uundaji na utathmini wa kamusi, ufundishaji lugha, uondoaji au upunguzaji wa utata wa kimaana katika matini, utoaji wa faharasa za Kingereza, uorodheshaji wa mizizi, mashina kama vidahizo, uchujaji

²¹Maagizo ya kutumia kongoo ya HCS :<http://www.aakkl.helsinki.fi/comeel/corpus/languagebank-instructions.pdf>

wa nomino za pekee katika matini (yaani majina ya watu, mahali n.k.), programu za tafsiri-mashine,

utabakishaji wa misamiati katika kongoo, utafiti wa idadi za utokeaji wa maneno fulani katika matini, uainishaji wa ujirani wa maneno, vielelezo vya upatanisho wa kisarufi, ufundishaji wa Kiswahili, uchambuzi wa misemo, nahau na methali na pia uhifadhi wa data za lugha miongoni mwa majukumu mengine.

Katika sura inayofuatia, data za Kiswahili zimewasilishwa kama mifano ya michakato hii ya kutumia programu za SALAMA katika kongoo ya HCS na kueleza zaidi mkao wa matini hizi .

3.7 Manufaa ya kongoo katika utafiti

Kwa mujibu wa Dash (2010:19-20), kongoo yenye sifa bainifu kama ilivyoelezwa hapo awali, basi huwa ni kipimo tosha cha kisayansi ambacho huwa kama kigezo cha kuyakinisha matumizi ya sifa mbalimbali za kiisimu zilizo katika lugha husika. Hii ni kwa sababu iwapo mtafiti atachanganua kanzi-data (data-base) ya kongoo anaweza kuibuka na habari tofauti kama;

i) Habari kuhusu lugha au lahaja zake kama vile: sauti, fonimu, lafudhi, herufi, uwakifishaji, mofimu, maneno, mizizi, shina, viini, virai, misemo, marudio rudio, methali, vishazi, sentensi na kadhalika, kuhusiana muundo, yaliyomo, mpangilio viambishi na minyambuliko, ruwaza za kupanga vyambajengo vya lugha, miktadha ya matumizi, ruwaza za matumizi, tofauti za miktadha na kadhalika.

ii) Habari za matumizi ya vipengee kama herufi, fonimu, mofimu, au sentensi zinazohusiana na matumizi ya kimaielezo, kijazanda, kiistaiara, kimsemo, au kifasihi na kadhalika.

iii) Habari za uziada-lugha zinazohusiana na muda, mahali, hali, mawakala wa matukio ya kilugha, mapisi na chimbuko la kijamii la matendo ya kilugha, diskosi mbalimbali na pragmatiki na pia maarifa ya maono ulimwengu ya watumiaji wa lugha husika kwa jumla.

iv) Habari za kiisimu na kiuamilifu za herufi, mofimu, virai, maneno , sentensi, misemo na nahu, methali na kadhalika kama vinavyohusiana na muundo, yaliyomo, mpangilio kutumia viambishi na minyambuliko, ruwaza za kupanga vyambajengo vya lugha, miktadha ya matumizi, ruwaza za matumizi, tofauti za miktadha hiyo na kadhalika.

Kongoo, kwa hivyo, inafaa kujengwa kwa msingi wa malengo maalum na umahsusi wa kazi yake. Kwa hivyo, kongoo hutofautiana kulingana na lengo la aliyeyunda, muundo wake na anayeitumia. Kompyuta huwa ni chombo muhimu sana kwa mwanaisimu-kongoo anapounda au kuchanganua kongoo iliyoundwa tayari. Uchakataji na uchanganuzi huu wa data iliyohifadhiwa huwa jukumu muhimu sana kwa mwanaisimu-kongoo. (Dash, 2010).

3.7.1 Matumizi mahsusi ya kongoo katika utafiti wa lugha

Kongoo yoyote inatumiwa moja kwa moja katika maeneo kama ufafanuzi wa lugha, uchunguzi wa

kisintaksia, kifonetiki na kifonolojia, lafudhi, mofolojia, arudhi, leksikolojia, semantiki, leksokografia, uchanganuzi wa diskosi, pragmatiki, kufundisha lugha, kupanga lugha, isimu jamii, Isimu-saiko, Semiotiki, Isimu-ng'amuzi (cognitive linguistics), na isimu kompyuta na kadhalika. (Dash, 2010).

Kupatikana kwa kongoo ambayo ni matini zinazosomeka kikompyuta kumefanya upatikanaji wa data kwa haraka na kwa urahisi uwepo na pia kuwasilisha data hiyo kwa muundo ufaao kwa uchanganuzi. Hili limetokana na kompyuta kuwezesha ukusanyaji, uhifadhi ma uchakataji wa kanzi-data za lugha asilia. Kwa hivyo, nyanja zilizotajwa hapo juu na kila eneo la lugha linaweza kushughulikiwa kikongoo (Clark.A.,Fox.C, na Lappin,S 2010., Dash, 2010., Hurskainen, 2005:122-123., Sinclair, J., 2004).

Kwa mujibu wa Dash, 2010:20, kongoo inaweza kushughulikia maeneo ya lugha yafuatayo;

i) Kongoo kama chanzo cha maarifa: Kongoo imetumika katika kuunda maktaba za lugha nyingi, kuunda vitabu vya viada vya kufundishia lugha, kuunda kamusi za lugha moja mtandaoni na katika machapisho, kuunda kamusi za lugha mbili katika mtandao na katika machapisho, kuunda thesauri za lugha moja katika mtandao na machapisho, kuunda kamusi ya kusomeka kikompyuta yaani (Machine Readable Dictionaries-MRD), kuundia hazina ya lugha anuwai ya misamiati na kuunda kamusi za kielekroniki zenye manufaa mengi kuliko za machapisho na kadhalika.

ii) Kongoo katika Teknolojia ya lugha: kongoo inatumika katika kusanifia zana na mifumo ya kikompyuta ya kuchakata neno, kukagua tahajia, kuhariri matini, kuchanganua mofolojia, kuchambua sentensi, kuhesabu takwimu katika matini k.m. neno hutokea mara ngapi katika matini , usakuzi wa vipashio vya lugha, kufupisha matini-kimuhtasari, kualamisha matini, kupata habari muhimu, kupatanisha kisarufi, kupunguza au kuondoa utata wa maana ya maneno na kadhalika. Zana na programuhizi ni kama, SWATWOL, SWACGP, SWASENT, WordNet (synset), semantic web, Semantic Net, uainishaji wa maneno katika lugha, ukundishaji wa maneno ya kawaida (*Local Word Grouping*), na kadhalika.

iii) Kongoo kwa usaidizi wa mifumo ya tafsiri-mashine: Kongoo hutumika katika kufikia mifumo ya hazina za lugha, mifumo ya tafsiri -mashine, kufikia mifumo ya lugha anuwai na kupata habari katika mifumo ya lugha moja hadi nyingine na kadhalika.

- iv) **Kongoo kwa mifumo ya mahusiano ya binadamu—kwa—maschine:** Kongoo hutumika kwa mifumo ya kutambua maandishi yanayoonekana yaani (*Optical Character Recognition-OCR*), kutambua sauti, kubadilisha matini hadi kuwa sauti, masomo ya mtandaoni, kufundisha mtandaoni, kuandaa matini mtandaoni, masomo ya lugha yanayoendeshwa kikompyuta, maagizo yanayoendeshwa kikompyuta, usimamizi wa kikompyuta na kadhalika.
- v) **Kongoo katika teknolojia ya sauti:** Kongoo sauti inatumika kutayarisha muundo mzima wa teknolojia ya sauti, fonetiki, msamiati na matamshi, kutambua msemaji kama kwenye simu, kurekebisha hali za ulema wa sauti na katika isimu ya kiupelelezi (kiforensiki).
- vi) **Kongoo katika Isimu ya kawaida:** hutumika katika ufafanuzi wa lugha, uchunguzi wa kisintaksia, kuandika vitabu vya sarufi, kifonetiki na kifonolojia, lafudhi, mofolojia, arudhi, leksikolojia, kuchunguza semantiki, kutafsiri lugha moja hadi nyingine, uundaji wa kamusi za lugha mbili, ujenzi wa hifadhi ya istilahi za lugha, leksikografia, uchanganuzi wa diskosi, pragmatiki, kufundisha lugha, uchunguzi wa sarufi, kupanga lugha, Isimu jamii, Isimu-Saiko, Semiotiki, Isimu-ng'amuzi (*Cognitive Linguistics*) na kadhalika.

3.8 Hitimisho

Sura hii imetalii nafasi ya kongoo matini katika teknolojia ya lugha kwa kumulika dhana nzima ya mitazamo ya kikongoo katika utafiti wa lugha kiteknolojia. Imedhihirika kwamba bila kongoo, taaluma ya Isimu-Kompyuta na michakato ya kufanyia utafiti lugha kiteknolojia haiwezi kukamilika. Kongoo, kwa hivyo, ndiyo mali-ghafi na pia uwanja wa kujaribia mbinu zote na programu zote za kikompyuta kwa kuwa kongoo si mkusanyiko wa matini tu ila ni lugha iliyowakilishwa na sehemu kidogo katika uhalisia wake. Kongoo ya HCS iliyotumika katika kujaribia programuza SALAMA na pia iliyoainishwa na programu za SALAMA kama SWATWOL inaonekana kitakwimu na kihithibati kuwa ni kongoo inayowakilisha lugha ya Kiswahili kikamilifu. Mitazamo ya kikongoo katika utafiti wa lugha kikompyuta imedhihirika kama nadharia tosha katika utafiti wa lugha asilia kwa ufasaha zaidi kwa kutumia kompyuta. Katika sura inayofuata ya uwasilishaji wa data, tumeangalia mifano ya ufaafu wa Kongoo ya HCS katika utafiti wa Kiswahili kwa kutumia teknolojia ya lugha.

SURA YA NNE

UCHAMBUZI NA UFAFANUZI WA DATA (TEKNOLOJIA YA LUGHA KATIKA UTAFITI WA KISWAHILI)

4.0. Utangulizi

Katika sura ya kwanza ya utafiti huu tumeweza kuweka msingi kuhusu hali ilivyo katika utafiti wa lugha kutumia teknolojia ya lugha. Sura ya pili ilijadili na kubainisha programu mbalimbali za kikompyuta na michakato iliyotumiwa na wataalamu walioshughulikia lugha ya Kiswahili katika mradi wa SALAMA pamoja na ujenzi wa mradi huo. Sura ya tatu nayo ilieleza kipengele cha kongoo ambacho ndio msingi wa lugha au ‘mali ghafi’ yanayotumiwa na programu hizi za kompyuta. Sura hii, kwa hivyo, ni muhimu katika kutoa ithibati ya kuwa mradi wa SALAMA ulitimia teknolojia ya lugha katika uchunguzi wa lugha ya Kiswahili na kuonyesha kuwa mbinu hiyo inaweza kutumika katika utafiti wa lugha hii. Sura hii kwa hivyo itajikita katika uwasilishaji na uchanganuzi wa data ambazo ni mifano mahususi na maelezo yanayodhihirisha wazi jinsi mradi wa SALAMA ilivyoshughulikia nyanja mbalimbali za lugha ya Kiswahili kwa kutumia programu za mbalimbali za kikompyuta zikiwemo zilizotajwa katika sura ya pili na nyingine zaidi. Maelezo yanayotokana na data za hojaji pia yameshirikishwa katika viwango mbalimbali vya kazi hii kwa mujibu wa malengo ya utafiti.

4.1 Uwasilishaji na Uchanganuzi wa data

Uteuzi wa mifano hii na uchanganuzi wake umeongozwa na nadharia zetu na malengo ya utafiti tulioweka. Uwasilishaji na uchanganuzi data huu utaoongozwa na mihimili mikuu ya nadharia mbili za utafiti wetu ambazo ni Nadharia ya Msambao wa Ugunduzi katika vigezo vyake vitano vya upenyaji wa ugunduzi ambavyo ni manufaa ya teknolojia mpya, Ufaafu wa kuendana na wakati, uwezekano wa kimatumizi na ushuhudiwaji au uathirianaji (Rogers, 2003) na pia Nadharia ya Kiisimu Kongoo (Nadharia wa Mtazamo wa Kikongoo) ambayo inasisitiza kuwa maana ya lugha hupatikana katika matumizi yake, kongoo matini ni ushahidi wa lugha halisi inayoonyesha kuwa muundo na maana vinahusiana (Sinclair 2004) na aidha kongoo huhitaji zana au programu (za kikompyuta) kuielezea lugha. Ni muhimu ieleweke kimbele kuwa baadhi ya mifano ya data hizi inaweza kutumika katika muktadha zaidi ya mmoja; kwa mfano ufafanuzi wa tungo-funge kama nahau unaweza kutumika katika tafsiri-mashine, uundaji kamusi au hata uchunguzi wa kisemantiki na kadhalika.

4.2 Teknolojia ya lugha katika utafiti wa Mofolojia ya Kiswahili

Kwa mujibu wa Katamba, F. na Stonham J., (2006), mofolojia ndilo tawi la isimu lililo katikati ya matawi mengine. Wasomi hawa wanatoa taswira ya uhusiano wa mofolojia na taaluma nyingine za isimu kama fonolojia, sintaksia, semantiki na kuonyesha jinsi uchanganuzi wa neno ulivyo muhimu katika kufanyia utafiti lugha yoyote hiyo. Mawazo haya ambayo yanayoungwa mkono na Spencer, A (1993) na Bubenik, V., (2003) wanaosisitiza kuwa kutokana kwa uchambuzi wa muundo wa ndani wa neno (Mofolojia) ndipo tunaielewa maana yake (semantiki) na kuweza kuchambua mpangilio wake katika sentensi (sintaksia), pia inatilia uzito uhalali wa matumizi ya nadharia ya mtazamo wa kongoo ambao unatilia maanani lugha katika uhalisia wake ambayo ni mkusanyiko wa maneno.

Kama tulivyotaja hapo mwanzoni mradi wa SALAMA umejengeka katika msingi wa uchanganuzi wa Kiswahili kimofolojia kabla ya vipengele vinginevyo vya kushughulikia lugha hii kuingizwa humo (Hurskainen 1992). Hii ndiyo programu ya kwanza ya SWATWOL iliyoanza kwa kuchanganua matini za Kiswahili kimofolojia ikifuatia utaratibu wa Koskenniemi, K. (1983) wa mofolojia ya tabaka-mbili. Baadaye kuliongezeka uwezo wa kuondoa utata wa maana ya matini chini ya mfumo wa Sarufi Kikwazo (SWACG), kuchambua Kiswahili Kisintaksia na hatimaye ikashirikisha kushughulikia matini ambazo si neno tu ila ni makundi ya maneno au tungo-funge kama tunavyoendelea kuziita humu (Hurskainen 2006). Hurskainen anasema kuwa kuna manufaa ya kuchanganua Kiswahili kimofolojia kutumia mbinu mpya ya kimofolojia ya ufafanuzi wa lugha kwa awamu mbili (SWATWOL mpya) badala ya mofolojia ya tabaka mbili na muundo-ukomo na baadaye katika muundo wa awamu mbili kama tuliyoieleza katika sura ya kwanza na ya pili.

Kwa mujibu wa Hurskainen 2004b, programu hii ya SWATWOL mpya inafanya ufafanuzi wa kimofolojia kwa kutumia mfumo wa awamu mbili na kufikia mwaka wa 2004 programu hii ilikuwa imepata uwezo wa kufafanua lugha ya Kiswahili sanifu kutoka kwa matini zozote za Kiswahili sanifu kwa ufasaha wa asilimia 100, kumaanisha kuwa neno la Kiswahili linapata maana yake sahihi katika muktadha lilivyotumiwa na pia katika miktadha mingine yote kwenye matini hiyo. Hiki ni kiwango cha juu sana kwa lugha ya Kiafrika ambayo haina historia ndefu ya kutafitiwa kwa kutumia teknolojia ya lugha kama za Ulaya mathalan Kiswahili.

Katika sehemu hii, tutaongozwa na pia kutumia vifupisho vifuatavyo katika data zetu zitakazowasilishwa na kuchanganuliwa;

‘...ADJ - kivumishi
A-INFL- Kivumishi kinachonyambulika
CAP – Neno lenye herufi kubwa

DEM –Kiwakilishi kionyeshi
 IDIOM-V> -kitenzi katika nahau
 <IDIOM – nahau au sehemu katika tungo lililo nahau
 MT – Tafsiri-mashine
 MWE – tungo-funge
 MW-N>> Nomino ikiwa katika tungo-funge na maneno mengine 2 kulia
 <<MW neno likiwa katika tungo-funge na maneno mengine 2 kushoto kwake
 <MW> neno likiwa katika tungo-funge na maneno mengine 2 mengine moja
 kushoto la jingine kulia
 N nomino
 OBJ kiambishi awali cha yambwa
 PAST- njeo iliyopita
 PERF:me – Kiwakilishi cha hali timilifu- *me*
 POSS Kiwakilishi cha kimilikishi
 PR:na – Njeo ya wakati ujao kiambishi- *na*
 PRON kiwakilishi
 REDUP –neno la kurudiwa rudiwa (k.m. mbalimbali)
 REL Kiambishi rejeshi
 SALAMA- Swahili Language Manager
 SP – kiambishi awali cha kiima
 SVO Kitenzi kielekezi
 TAM- Kiambishi cha tense njeo na dhamira
 VE –Kiambishi tamati cha kitenzi
 VFIN Hali ya kitenzi-ukomo
 @FMAINVtr+OBJ> Kitenzi ukomo kikuu kielekezi chenye yambwa bainifu/wazi
 @FMAINVtr-OBJ> Kitenzi ukomo kikuu kielekezi kisicho na yambwa bainifu/wazi
 @OBJ –Yambwa ya kishazi
 @SUBJ Kiima katika kishazi...” (Hurskainen 2009:1)

NB: ‘z’ ni kipashio cha muda cha kualamisha maelezo ya kitenzi na iwapo kanuni itatumika kama ‘:z’, ni ili kanuni nyingine isibadilishe maelezo hayo.(Hurskainen 2009).

Ni kweli kuwa unaweza kushughulikia kikompyuta matini za kawaida za Kiswahili kama alivyofaulu kufanya katika SALAMA (Hurskainen 2004a) ziwe ni katika vitabu, magazeti, misahafu au hata kutoka mtandaoni bora matini yenyewe iwe katika mfumo wa kusomeka kikompyuta.

4.2.1 Uchanganuzi wa kimofolojia kutumia SWATWOL

Katika sehemu hii tutatoa na kueleza mifano michache ambayo itatufafanulia hatua zilizotumika na Hurskainen 2004, 2006, 2008 na 2009) kuchanganua matini za Kiswahili kutumia SALAMA. Kusudi lilikuwa ni kupata kila aina ya neno katika msamiati wa lugha hiyo likiwa na ufafanuzi wa kina wa kimofolojia.

MFANO 1.

‘...anayekisoma
 "soma" V 1-SG3-SP VFIN { he } PR:na 1-SG3-REL { who } 7-SG-OBJ OBJ
 { it } z [soma] { read } SVO...’

Kutoka: Ripoti Nambari 2 ya 2009:6

Ukiangalia mofimu zilizo katika (neno) kitenzi hiki utaona zimechanganuliwa vifuatavyo
 a=kiambishi cha kiima cha nafsi ya tatu umoja ngeli ya 1. Kikompyuta inakuwa (1-SG3-SP)
 na=kiambishi cha njeo –wakati uliopo ((PR:na)

ye- kirejeshi cha nafsi ya tatu umoja, ngeli ya 1 (1-SG3-REL)
ki= yambwa tendewa ya ngeli ya 7 umoja (7-SG-OBJ)
som= mzizi wa kitenzi
a=kiambishi tamati cha kauli ya kitenzi.

i) Awamu ya kwanza

Utatambua kuwa katika mfano 1, kuna baadhi ya mambo yamebakia nje kama shina la kitenzi ‘soma’ ni (POS) aina ya neno iitwayo kitenzi (V), ni kitenzi ukomo (VFIN) nan i kielekezi (SVO). Kwa kuwa haya pia yatahitajika katika kuainisha neno hilo, basi kuna haja ya uchanganuzi mwingine wa kimofolojia wa kina zaidi ambao utafanywa katika awamu ya pili baada ya huu wa juujuu (Hurskainen 2009).

MFANO 2.

‘...anayekisoma
"soma" [soma] V aSP naTAM yeREL kiOBJ z aVE REDUP...’

MFANO 3

‘...anayekisomasoma
"soma" [soma] V aSP naTAM yeREL kiOBJ z aVE somaREDUP...’

Kutoka: Ripoti Nambari 2 ya 2008:4

ii) Awamu ya pili

Katika hatua hii tunashuhudia upanuzi wa viwango vya uchanganuzi ili habari zijitokeze kwa kina na kila mofimu kuelezewa tofauti na awamu ya kwanza iliyotoa tu kategoria za kisarufi za maneno za vpashio vya kimofolojia. Tazama mfano wa 4 ulivyofafanua viambishi vya ngeli za nomino, nafsi, umoja na wingi na aina ya kitenzi VFIN- kitenzi ukomo.

MFANO 4

anayekisoma
"soma" [soma] V 1-SG3-SP VFIN PR:na 1-SG-REL 7-SG-OBJ z

Kutoka:Ripoti Nambari 2 ya 2008:4

Tuangalie jinsi SALAMA ilivyoshughulikia hali inapokuwa na utata wa maana katika aina ya kitenzi kama ‘**unaouleta**’.

MFANO 5

unaouleta
"leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 11-SG-REL { which }
11-SG-OBJ { it } z { bring }

Kutoka: Ripoti Nambari 2 ya Hurskainen 2009:4

MFANO 5

Hatua ya kwanza ya uainishaji wa kitenzi inatoa maelezo sahili ila vipengele baadhia havijafanuliwa kwa kina kwa mfano hatufahamu kitenzi ni cha aina gani na yambwa ni ya ngeli gani au ni katika umoja au wingi.

unaouleta
"leta" [leta] V uSP naTAM oREL uOBJ z aVE REDUP

Kutoka: Ripoti Nambari 2 ya (Hurskainen 2009:5)

MFANO 6

Hatua hii ya pili ya uainishaji wa kitenzi kile kile kimofolojia inaibuka na njia 18 ambazo zote ni sahihi kisarafu. Mfano wa sita unaonyesha sababu za utata huu wa maana ni kuwa viambishi vya kiima na vya kirejeshi vina maana tatu tatu tofauti ila sahihi na kiambishi cha yambwa nacho kina fasiri 2 tofauti sahihi. Hili linadhihirisha kuwa, ili kuweza kuchunguzi lugha yoyote kwa ufasaha ni muhimu kila neno lifanyiwe uainishaji wa kina wa kimofolojia. (Hurskainen 2009).

unaouleta

"leta" [leta] V 1-SG2-SP VFIN PR:na 2-PL-REL 3-SG-OBJ z
"leta" [leta] V 1-SG2-SP VFIN PR:na 2-PL-REL 11-SG-OBJ z
"leta" [leta] V 1-SG2-SP VFIN PR:na 3-SG-REL 3-SG-OBJ z
"leta" [leta] V 1-SG2-SP VFIN PR:na 3-SG-REL 11-SG-OBJ z
"leta" [leta] V 1-SG2-SP VFIN PR:na 11-SG-REL 3-SG-OBJ z
"leta" [leta] V 1-SG2-SP VFIN PR:na 11-SG-REL 11-SG-OBJ z
"leta" [leta] V 3-SG-SP VFIN PR:na 2-PL-REL 3-SG-OBJ z
"leta" [leta] V 3-SG-SP VFIN PR:na 2-PL-REL 11-SG-OBJ z
"leta" [leta] V 3-SG-SP VFIN PR:na 3-SG-REL 3-SG-OBJ z
"leta" [leta] V 3-SG-SP VFIN PR:na 3-SG-REL 11-SG-OBJ z
"leta" [leta] V 3-SG-SP VFIN PR:na 11-SG-REL 3-SG-OBJ z
"leta" [leta] V 3-SG-SP VFIN PR:na 11-SG-REL 11-SG-OBJ z
"leta" [leta] V 11-SG-SP VFIN PR:na 2-PL-REL 3-SG-OBJ z
"leta" [leta] V 11-SG-SP VFIN PR:na 2-PL-REL 11-SG-OBJ z
"leta" [leta] V 11-SG-SP VFIN PR:na 3-SG-REL 3-SG-OBJ z
"leta" [leta] V 11-SG-SP VFIN PR:na 3-SG-REL 11-SG-OBJ z
"leta" [leta] V 11-SG-SP VFIN PR:na 11-SG-REL 3-SG-OBJ z
"leta" [leta] V 11-SG-SP VFIN PR:na 11-SG-REL 11-SG-OBJ z

Kutoka: Ripoti Nambari 2 ya 2009:5

4.2.2 Kushirikisha habari za Kisemantiki katika uainishaji wa kimofolojia

MFANO 7

SALAMA imejaribu hatua kuongezea habari za kisemantiki na kileksika katika uainishaji huo baada ya kupata msamiati uliainishwa kimofolojia na kuongezewa maelezo ya Kiingereza kwa kila mofimu. Mfano huu wa 7 unajumuisha hata shina la kitenzi.

‘...*unaouleta*

"leta" V [leta] SVO 1/2-SG2-SP VFIN { you } PR:na 1/2-PL-REL { who } 3/4-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 1/2-SG2-SP VFIN { you } PR:na 1/2-PL-REL { who } 11-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 1/2-SG2-SP VFIN { you } PR:na 3/4-SG-REL { which } 3/4-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 1/2-SG2-SP VFIN { you } PR:na 3/4-SG-REL { which } 11-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 1/2-SG2-SP VFIN { you } PR:na 11-SG-REL { which } 3/4-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 1/2-SG2-SP VFIN { you } PR:na 11-SG-REL { which } 11-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 3/4-SG-SP VFIN { it } PR:na 1/2-PL-REL { who } 3/4-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 3/4-SG-SP VFIN { it } PR:na 1/2-PL-REL { who } 11-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 3/4-SG-SP VFIN { it } PR:na 3/4-SG-REL { which } 3/4-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 3/4-SG-SP VFIN { it } PR:na 3/4-SG-REL { which } 11-SG-OBJ { it } z { bring }
"leta" V [leta] SVO 3/4-SG-SP VFIN { it } PR:na 11-SG-REL { which } 3/4-SG-OBJ { it } z { bring }

"leta" V [leta] SVO 3/4-SG-SP VFIN { it } PR:na 11-SG-REL { which } 11-SG-OBJ { it } z { bring }
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 1/2-PL-REL { who } 3/4-SG-OBJ { it } z { bring }
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 1/2-PL-REL { who } 11-SG-OBJ { it } z { bring }
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 3/4-SG-REL { which } 3/4-SG-OBJ { it } z { bring }
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 3/4-SG-REL { which } 11-SG-OBJ { it } z { bring }
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 11-SG-REL { which } 3/4-SG-OBJ { it } z { bring }
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 11-SG-REL { which } 11-SG-OBJ { it } z { bring }...

Kutoka: Ripoti Nambari 2 ya 2009:6

Data hii ya mfano wa 7 inaonyesha kuwa neno *unaouleta* limeainishwa kikamilifu katika hatua hii. Uainishaji huu hata unaweza kuibuka na kanuni za kuondoa utata (jambo ambalo tutalimulika katika kipashio cha kisemantiki) ila muhimu kataja kuwa kanuni za kuondoa utata katika lugha huegemea sana muktadha. Asemavyo (Hurskainen 2009), alipotumia kongoo kupanua mazingira ya neno hilo aliibuka na muktadha ufuatayo ambayo tutaitathmini katika mwisho wa kijishemu hiki. *Umoja unaouleta uhuru*, (11). Utata wa maana sasa utaondolewa katika mfano wa 8 unaoeleza muktadha na kinachorejelewa.

MFANO 8

'... *umoja
 "umoja" N 11-SG { unity }
 Unaouleta
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 11-SG-REL { which } it } z { bring }
 Uhuru
 "uhuru" N 11-SG { freedom }

'", PUNC { , } ...'

Kutoka: Ripoti Nambari 2 ya 2009:6

4.2.3 Kushirikisha uchambuzi wa kisintaksia katika uainishaji wa kimofolojia

MFANO 9

SALAMA iliweza kupiga hatua zaidi kwa kutoa uainishaji wa kisintaksia wa neno *unaouleta* kwa kutumia mbinu ya sarufi kikwazo ya SWACG; kama ilivyo katika mfano huu. Alama za kisintaksia zina @kama kiambishi awali kwa hivyo @SUBJ na @OBJ huwa kiambishi kiima na kiambishi cha yambwa kwa mfuatano huo na kitenzi huwa (@FMAINVtr+OBJ>) kumaanisha ni kitenzi kikuu cha ukomo kina kiambishi awali cha yambwa na ni kielekezi na kina yambwa moja bainifu upande wa kulia.

*umoja

"umoja" N 11-SG { unity } @SUBJ
 unaouleta
 "leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 11-SG-REL { which } 11-SG-OBJ { it } z { bring } @FMAINVtr+OBJ>
 uhuru
 "uhuru" N 11-SG { freedom } @OBJ
 ,

"," PUNC { , }

Kutoka: Ripoti Nambari 2 ya 2009:6

MFANO 10

Mfano huu unatuonyesha jinsi uainishaji huo wa kisintaksia unavyokuwa iwapo kutakuwa kumekosekana yambwa bainifu na hivyo ualamishaji wa kitenzi utakuwa (*@FMAINVtr-OBJ*>)

```
*umoja
"umoja" N 11-SG { unity } @SUBJ
unaouleta
"leta" V [leta] SVO 11-SG-SP VFIN { it } PR:na 11-SG-REL { which }
11-SG-OBJ { it } z { bring } @FMAINVtr-OBJ>
?
"," PUNC { , }
```

Kutoka: Ripoti Nambari 2 ya 2009:6

4.2.4 Kuelezea tungo-funge kimofolojia katika tafsiri

Kama tulivyotaja hapo awali baadhi ya dhana hizi huingiliana pakubwa na pia uainishaji wa msamiati wa Kiswahili kimofolojia huwa ndiyo msingi wa mbinu za SALAMA katika utafiti wa lugha hii. Tutatoa mifano zaidi kuhusu tungo-funge na changamoto zake pamoja na umahsusi wake katika teknolojia ya lugha hasa katika tafsiri-mashine na uundaji wa kamusi kikompyuta, ila ni vyema kuzitolea mifano michache katika muktadha huu wa uainishaji wa kimofolojia. Hii ni kwa sababu kundi la maneno linaweza kuwa tungo-funge katika mazingira fulani na kwingineko yakawa ni mfuatano wa maneno tu kama asemavyo (Hursakainen 2009:7) ambaye anapendekeza na alitumia mbinu ya hatua kwa hatua ya Sarufi Kikwazo ya SWACG na anaiona bora zaidi licha ya kuwa MWE zinaweza kutengwa kutumia uainishaji wa kawaida. Mfano ufuatao unashughulikia tungo-funge ambayo nahau yenye kitenzi na yambwa.

MFANO 11

```
*alipiga
"piga" V [piga] SVO ACT 1/2-SG3-SP VFIN { he/she } PAST z { hit }
picha
"picha" N ENG 9/10-SG { picture }
```

MFANO 12

```
*alipiga
"piga" V [piga] SVO ACT 1/2-SG3-SP VFIN { he/she } PAST z { hit }
IDIOM-V>3
picha
"picha" <IDIOM { photograph }
```

Kutoka: Ripoti Nambari 2 ya 2009:6

MFANO 13

Katika awamu ya pili, maelezo ya Kiingereza ya kileksika hutolewa kutoka kwa kitenzi na ufafanuzi wa mwisho unabakia kwenye neno la mwisho ifuatavyo;

```
*alipiga
"piga" V [piga] SVO ACT 1/2-SG3-SP VFIN { he/she } PAST z IDIOM-V>
@FMAINVtr-OBJ>
picha
"picha" <IDIOM { photograph }
```

Kutoka: Ripoti Nambari 2 ya 2009:6

MFANO 14

Sehemu yote hatimaye inafungwa kwa mabano baada ya kuondoa lugha asilia kwa kuwa Kiswahili kwa sasa hakihitajiki katika tafsiri hii.

(V SVO ACT 1/2-SG3-SP VFIN {he/she} PAST z IDIOM-V> @FMAINVtr-OBJ>
<IDIOM {photograph})

Kutoka: Ripoti Nambari 2 ya 2009:8

MFANO 15

Tungo-funge zinaweza hata kuwa na zaidi ya maneno mawili kama hii '*umoja wa mataifa*' yenye maneno matatu

'... *umoja
"umoja" N 11-SG { unity }
wa
"wa" GEN-CON 11-SG x { of }
*mataifa
"taifa" N AR 5/6-PL { nation }...'

Kutoka: Ripoti Nambari 2 ya 2009:9

MFANO 16

Kwa kuwa hii imezoeleka sana kimatuzi, haiwezi kutafsiriwa katika muundo wa lugha pokezi (Kiingereza) kwa hivyo SALAMA inatumiwa kuialamisha kama MWE (tungo-funge) na kisha kuweka maelezo mapya katika neno la mwisho la tungo-funge hilo kama ilivyo hapa chini.

*umoja
"umoja" N 11-SG { unity } MW-N>>4
wa
"wa" MW<>
*mataifa
"taifa" <<MW { *united *nations }

Kutoka: Ripoti Nambari 2 ya 2009:9

MFANO 17

Hatimaye maneno ambayo hayana kazi yanaondolewa na tungo-funge inafungiwa kwa mabano kama katika mfano huu;

(N 11-SG MW-N>> <<MW { *united *nations }) 5

Kutoka: Ripoti Nambari 2 ya 2009:11

4.2.5 Kudhibiti Mpangilio wa maneno

Kirai nomino cha Kiswahili kinatofautiana na cha Kiiingereza kwa jinsi maneno yanavyopangiliwa. Katika Kiingereza nomino hufuata maneno mengine ila Kiswahili ni tofauti kwa kuwa nomino hutangulia maneno mengine kama ifuatavyo (Hurskainen 2009).

MFANO 18

*watoto
"mtoto" N 1/2-PL { the } { child } CAP
wangu
"angu" PRON POSS 1/2-PL SG1 { my }
wazuri
"zuri" ADJ A-INFL 1/2-PL { good }
hawa
"hawa" PRON DEM :hV 1/2-PL { these }

wamesoma

"soma" V 1/2-PL3-SP VFIN NO-SP-GLOSS PERF:me z [soma] { read } SVO

Kutoka: Ripoti Nambari 2 ya 2009:11

MFANO 19

Kirai nomino kinawekwa kwa msatri wake kila neno likifungiwa kwa mabano.

(N 1/2-PL { the } { child } CAP) (PRON POSS 1/2-PL SG1 { my }) (ADJ
A-INFL 1/2-PL { good }) (PRON DEM :hV 1/2-PL { these }) (V 1/2-PL3-SP
VFIN NO-SP-GLOSS PERF:me z { read } SVO)

Kutoka: Ripoti Nambari 2 ya 2009:11

MFANO 20

Hapa inabidi kupangua mpangilio wa maneno ili uingiliane na wa Kiingereza

(PRON DEM :hV 1/2-PL { these }) (PRON POSS 1/2-PL SG1 { my }) (ADJ
A-INFL 1/2-PL { good }) (N 1/2-PL { child } CAP) (V 1/2-PL3-SP VFIN
NO-SP-GLOSS PERF:me z { read } SVO)

Kwa hivyo itabakia hatimaye; *these my good children have read.*

Kutoka: Ripoti Nambari 2 ya 2009:11

4.2.6 Uchanganuzi wa data

Katika kuichunguza lugha ya Kiswahili kimofolojia imedhihirika kuwa SALAMA imetekeleza majukumu kadha kupitia programu zake zinazoainisha msamiati wa lugha katika kongoo matini. Ni dhahiri kwa mujibu wa data hii kuwa uondoaji wa utata wa maana unasaidia ufafanuzi wa kimofolojia katika Kiswahili. Mbinu iliyotumiwa katika SWATWOL mpya; ambayo ni uainishaji wa kimofolojia wa awamu-mbili imefaa zaidi ikilinganishwa na mbinu za zamani tulizotaja katika sura ya pili. Ni dhahiri kuwa utafiti wa kimofolojia unaathiri maeneo mengi ya kuitafitia lugha kwa kuwa sintaksia, semantiki, tafsiri na ushughulikiaji wa tungo-funge umeweza kushughulikiwa na SWATWOL.

Natija nyingine inayopatikana katika uainishaji huu wa kimofolojia kutumia teknolojia ya lugha ya SALAMA ni sambamba na nadharia ya Msambao wa Ugunduzi kipengele cha manufaa kwa mtumiaji na kile cha ufaafu au kuendana na wakati, ni kuwa kama asemavyo Hurskainen 2009, zana za kikompyuta zinazotumika katika uainishaji wa awamu ya pili baada ya ile ya kijuu juu, ni programu iitwayo Beta. Beta ni programu inayopatikana bila malipo na kwa hivyo hata kuongeza thamani zaidi mchakato wa teknolojia ya lugha kwa kuwa aina gharama kwa watumiaji wake, na matokeo yake yameonekana kufidia pengo la utafiti wa lugha kimofolojia katika tafsiri ya makundi ya maneo (tungo-funge) kama ilivyo katika mfano wa 10, 11, 12, 13, na 14.

Katika mfano wa 15 pia tunaona jinsi programu hii inavyosaidia katika tafsiri ya tungo-funge zilizozoeleka kama Umoja wa Mataifa. Hili linamaanisha kuwa teknolojia ya lugha ikitumia

kongoo inaweza kukumbuka maneno yenye umashuhuri fulani na hivyo kupunguza udhia wa kuunda kanuni nyingi za kikompyuta katika uainishaji wa istilahi hizo kwa kuwa kompyuta itayakumbuka kwa urahisi na kufanya mkato katika uwasilishaji wake kimofolojia.

Data hii pia inaonyesha jinsi mtazamo wa kikongoo unavyofaidi utafiti wa teknolojia ya lugha kwa kuwa mathalan, katika Mfano wa 6 na 7, uongezaji wa habari za kisintaksia na kisemantiki na hatimaye, kuingiza msamiati ulioainishwa katika kongoo ili kuupa muktadha, kunaibua udhahiri wa utafiti kwa kufungua mawanda ya maana kwa muda mfupi kuliko vile tunganidhihirisha kwa njia za kiungamuzi au kufikiria mifano na miktadha hiyo.

Katika mfano wa 6, uainishaji wa kina wa kimofolojia, umeweza kuibuka na njia 18 ambazo zote ni maelezo sahihi kisarufi. Uwezo kama huu wa kiteknolojia ni muhimu sana katika utafiti wa lugha kwa kuwa mtafiti atafahamu miktadha na mitazamo kadha ya kuainisha vipashio fulani vya lugha na hatimaye, kumakinikia atakacho bila kupoteza muda kwa kubofya kifute na kuingia katika hatua nyingine ya kikompyuta.

Kwa mujibu Ng'ang'a 2005 matokeo kama haya ya uchanganuzi wa kimofolojia na uondoaji wa utata wa maana ya maneno ndiyo husaidia katika uundaji wa programu za kufitia lugha kikompyuta kama vile vikagua-sarufi, viorodheshi kwa vistari na pia mfumo kama ule aliotumia Sewangi (2001) wa kuunda istilahi za Kiswahili za sajili na nyanja mahsusi ya lugha kama kama fasihi na matibabu na kadhalika.

4.3 Teknolojia ya lugha katika utafiti wa Semantiki ya Kiswahili

Kama tulivyotanguliza kusema katika sura ni kuwa sehemu tofauti za lugha huingiliana. Tayari inakumbukwa kuwa Mfano wa 7 na wa 8 hapo juu, licha ya kuwa msingi wa uchanganuzi wake ulikuwa wa kimofolojia, mchakato wa kifikia katika mifano hiyo ulijihusisha na uchunguzi wa kisemantiki pia. Maana katika maneno huwa ni muhimu sana katika utafiti wa lugha yeyote. Iwapo maana lengwa haijaleweka, basi jukumu la lugha la mawasiliano huwa halijatimizwa. Hapa ndipo kongoo huwa muhimu kwa kuwa lugha halisi hupatikana katika mfumo unaoweza kutambulika kikompyuta na kuchunguzwa jinsi inavyotoa maana lengwa katika muktadha wa matumizi ya vipashio vyake mbalimbali vikiwemo, maneno, tungo-funge, sentensi na kadhalika.

Kwa mantiki hii basi, utata wowote wa maana katika lugha huwa ni kikwazo kinachohitaji kiondolewa ndipo maana lengwa ijitokeze. Utafiti wa semantiki ya Kiswahili katika mradi wa

SALAMA umesaidiwa sana na programu za uainishaji wa maneno kimofolojia kama tulivyobainisha hapo juu. Katika kijisehemu hiki tutawasilisha data yaani ithibati kupitia kwa mifano na ufafanuzi jinsi mradi wa SALAMA ulitumia pia programu za kuondoa au kupunguza utata wa maana kama zinazotumia kanuni za kuchagua kimuktadha, sarufi kikwazo pamoja na zile za ukisiaji au majaribio tulizofafanua katika sura ya pili katika kuchunguza Kiswahili kisemantiki.

4.3.1 Utata wa maana na Utatuzi wake katika Maneno ya Kiswahili

Ufasiri wa maana ya maneno huwa aghalabu si dhahiri kila wakati kwa kuwa neno moja huwa na zaidi ya maana moja na hivyo ufasiri huo kusemekana kuwa una utata wa maana. Mfano ufuatao unatoa data ya jinsi (Hurskainen, 2004) alitumia teknolojia ya lugha katika mradi wa SALAMA kushughulikia utata huo wa maana katika mchakato tutakaoendelea kutambua kama uondoaji wa utata wa maana. Katika mfano huu maneno yafuatao yanashughulikiwa katika kongoo ya HCS; *mwalimu, nyumbu, mbuzi, nta na taa*.

MFANO 21:

```

'... "<mwalimu>"
  "mwalimu" N 1/2-SG { teacher }

  "<nyumbu>"
  "nyumbu" N 9/10-0-SG { hartebeest } AN
  "nyumbu" N 9/10-0-PL { hartebeest } AN
  "nyumbu" N 9/10-0-SG { mule } AN
  "nyumbu" N 9/10-0-PL { mule } AN

  "<mbuzi>"
  "mbuzi" N 9/10-0-SG { goat } AN
  "mbuzi" N 9/10-0-SG { coconut grater }
  "mbuzi" N 9/10-0-PL { goat } AN
  "mbuzi" N 9/10-0-PL { coconut grater }

  "<nta>"
  "nta" N 9/10-0-SG { wax }
  "nta" N 9/10-0-PL { wax }

  "<taa>"
  "taa" N 5a/6-SG { large flat fish , skate } AN
  "taa" N 9/10-0-SG { lamp , lantern } AR
  "taa" N 9/10-0-SG { discipline , obedience }
  "taa" N 9/10-0-SG { large flat fish , skate } AN
  "taa" N 9/10-0-PL { lamp , lantern } AR
  "taa" N 9/10-0-PL { discipline , obedience }
  "taa" N 9/10-0-PL { large flat fish , skate } AN
  "utaa" N 11/10-PL { storage }
  "taa" ADJ A-UNINFL { exalted } AR ...' (Hurskainen, 2004a:369-370)

```

Data iliyo katika mfano wa 21 inaonyesha kuwa isipokuwa neno *mwalimu*, ambalo maana yake ni wazi, maneno mengine yote yana zaidi ya maana moja kwa hivyo yanasemekana kuwa yana utata

ya maana. Mathalan neno ‘nyumbu’ linazua utata wa kisemantiki kwa kuwa linarelejea nomino mbili mbili tofauti ambazo zote ni wanyama. Neno ‘mbuzi’ laweza kuwa mnyama au kifaa cha kukunia nazi, nalo ‘taa’ lina fasiri tisa au hata zaidi. Kwa hivyo tunahitaji kuteua maana ifaayo zaidi kwa msingi wa mkutadha wa matumizi, dhana tunayoita utatuzi au uondoaji wa utata wa maana (Hurskainen, 2004a).

MFANO 22:

Katika mfano ufuatao tutaona jinsi tunavyoweza kutatua utata wa maana ya neno kutumia kanuni ya kuchagua. Kiswahili na hali yake ya utapanisho wa kisarufi hutoa nafasi nzuri sana ya kutatua hata hali changamano za utata kwa kutumia kanuni za kuchagua (Hurskainen, 2004a) kama katika sentensi hii yenye utata; *‘Taa zote mbili zilivunjika’*

```

“... "<*taa>"
"taa" N CAP 5a/6-SG { large flat fish , skate } AN
"taa" N CAP 9/10-0-SG { lamp , lantern } AR
"taa" N CAP 9/10-0-SG { discipline , obedience }
"taa" N CAP 9/10-0-SG { large flat fish , skate } AN
"taa" N CAP 9/10-0-PL { lamp , lantern } AR
"taa" N CAP 9/10-0-PL { discipline , obedience }
"taa" N CAP 9/10-0-PL { large flat fish , skate } AN
"utaa" N CAP 11/10-PL { storage }
"taa" CAP ADJ A-UNINFL { exalted } AR
"<zote>"
"ote" PRON :ote 9/10-PL { all }
"<mbili>"
"mbili" NUM NUM-INFL CARD 9/10-PL { two }
"<zilivunjika>"
"vunjika" V 9/10-PL-SP VFIN { they } PAST z [vunja] { break , damage , annul }
SVO EXT: STAT :EXT
"<.$>"
"." { .$ }..."

```

(Hurskainen, 2004a:369-370)

Katika mfano huu tunashuhudia kuwa kuna utata katika neno ‘taa’ kwenye sentensi hii kama pale juu kwenye mfano 21 ila hakuna utata katika maneno mengine “zote”, “mbili” na “zilivunjika”. Kwa kuongozwa na sheria ya kisarufi ya kiambishi cha kiima, Hurskainen 2004a:372-373 anachunguza hali hii kwa kutunga sheria inayotumia alama za kimofolojia ambazo tayari zipo kwenye uainisho. Kwa mfano VFIN katika “zilivunjika” inaonyesha kuwa ni kitenzi ukomo kinachohitaji kiima upande wa kushoto na hivyo akaitungia sheria ya kwanza ya kuelezea kiima bila kuwazia ‘taa’ moja kwa moja kuwa ni kiima. Kanuni hiyo pia inaweza kuandikwa kama Mfano 24.

MFANO 23:

'...SELECT (N 9/10-0-PL lamp)
IF (1C (PRON))
(2C (NUM))
(3 (V) + (9/10-PL-SP)); ...'

(Hurskainen, 2004a:372)

MFANO 24:

"...SELECT (N 9/10-0-PL lamp)
IF (1C PRON)
(2C NUM)
(3 V + SP-10); ..."

(Hurskainen, 2004a:373)

Kanuni hii inamaanisha –chagua nomino iliyo katika maelezo ya Kiingereza iwapo neno linalofuatia ni kiwakilishi cha nomino na la pili kulia ni idadi na neno la tatu kulia ni kitenzi chenye kiambishi cha ngeli ya 10 (yaani 9/10-PL-SP). Tayari tumetatua utata wa neno 'taa' kikamilifu hata kama kanuni iliyoundwa ni mahsusi kwa muktadha huu.

Hata hivyo alama za uanishaji wa kimofolojia zinawezakutumika kuunda kanuni za kijumla ambazo zinatumika kwa upana zaidi. Hurskainen, 2004a anatunga kanuni hii ya 25 na 26 ambayo itazidi kutatua utata wa kiwango cha kisemantiki wa neno 'taa' zaidi.

MFANO 25:

SELECT (N 9/10-0-PL lamp)
IF (*1 SP-10 BARRIER VFIN)
(NOT *1 N BARRIER VFIN);

(Hurskainen, 2004a:373)

Katika mifano wa 24 na 25, uondoaji wa uatata wa maana umefaulu katambua maana lengwa ya 'taa' kutoka na kanuni za jumla kwa kutumia mfumo wa kuchagua na pia ule wa kikwazo(kuzuia/kupinga). Kwa mfano kanuni ya 25 inamzuia mtafiti asijisumbue kuendelea kutafuta zaidi iwapo kuna nomino yenye maana ya 'lamp' kwa Kiingereza iwapo upande wa kulia kuna kiambishi cha ngeli ya 10 na pia usiruhusu nomino yoyote kutokea kati ya neno lengwa na kitenzi ukomo. (Ingetokea basi utata wa 'taa' ungebakia bila kutatuliwa.

Kanuni zilizo katika mifano 24-25 zote zikitumika matokeo yatakuwa kama sentensi hii;

MFANO 27

'... "<*taa>"
"taa" N CAP 9/10-0-PL { lamp , lantern } AR
"<zote>"
"ote" PRON :ote 9/10-PL { all }
"<mbili>"
"mbili" NUM NUM-INFL CARD 9/10-PL { two }
"<zilivunjika>"
"vunjika" V 9/10-PL-SP VFIN { they } PAST z [vunja] { break , damage , annul } SVO
EXT: STAT :EXT
"<.\$>"
"." { .\$ } ..."

(Hurskainen, 2004a:373)

Mchakato huu wa utatuzi wa utata wa maana ya maneno katika mifano hii unaonyesha jinsi mbinu au kanuni za kuchagua huongozwa na sheria za kisarufi na zinavyofanikiwa kwa kuwa huzingatia ufafanuzi wa lugha kisarufi na pia hazihitaji mbinu zaidi za utatuzi. Kuna pia kanuni za kuondoa au Sarufi Kikwazo kama tulivyoziita tulipoelezea programu ya *SWACG* inazotumika wakati zile za kuchagua zinaposhindwa. Katika mifano ya 24 na 25, vikwazo vinavyotumika katika upendoaji wa utata ni kile cha kuelekeza nafasi ya uainisho katika tungo ili kanuni ifanye kazi na kwa mujibu wa Hurskainen 2004a:373, tunatambua jinsi kanuni inaruhusu au kuzuia kuenda mbele au kurudi nyuma hatua fulani au hadi upate alama fulani ya uainisho wa kimofolojia. Katika *SALAMA* matumizi ya *SWACG*, tunaona katika mfano wa 24 na 25 kuwa kikwazo kinaweza kuwa kinaruhusu au kinakataza sifa fulani ya kuwa kwenye neno au sentensi ili kuondoa utata wa maana za maneno. Kwa mujibu wa Ng'ang'a 2005:37, mfumo huu wa kuondoa utata uliotumiwa na Hurskainen katika *SALAMA* ulifanikiwa kuondoa utata wa maana za maneno ya Kiswahili kwa 94% na inaposhughulikiwa zaidi na kitatuzi cha kimajaribio cha *SWA-GUESS* inafikia hata 97% katika viwango vya kukabiliana na visa vya utata wa maneno yaliyobakia.

4.3.2 Uondoaji wa utata wa kisemantiki kutumia tafsiri-mashine na programu mseto

Katika tafiti bainifu zilizowahi kufanywa katika uwanja wa teknolojia ya lugha ya Kiswahili ni kazi ya Ng'anga' 2005 ambayo ilionyesha ukomavu mkubwa wa kushughulikia lugha kwa programu mseto za kikompyuta, hali iliyoonekana kama ni changamano sana ila ilizaa matunda kwa kupitia utaratibu wa hatua kwa hatua.

Kazi hii ambayo ilishirikisha baadhi ya programu na hazina za *SALAMA* kama kongoo ya HCS hasa Kamusi ya TUKI ililenga kufanya uchunguzi wa kiutaratibu wa aina na viwango vya utata wa kisemantiki katika Kiswahili kwa kutumia jukwaa la Kiingereza na pia kuibuka na mbinu mwafaka za kuondoa utata wa maana ya maneno ya Kiswahili kwa kutumia muktadha wa tafsiri-mashine. Kwa kutekeleza haya alitumia mbinu kadha ikiwemo ili ya kuweka nomino za Kiswahili katika makundi mahsusi na baadaye kutegemea mbinu za kikompyuta za kujifundisha lugha ambazo zinasaidia kuhifadhi uwezo wa kuondoa utata kwa kutumia habari inyopatikana kutoka kwa kongoo ya Kiswahili, kamusi ya Kiswahili kwa Kiingereza na programu huria zinazopatikana za Kiingereza kama WordNet.

Ili kufikia hatua ya uondoaji wa utata huu wa kisemantiki, (Ng'ang'a 2005) anatumia mbinu za kikongoo ambapo mfumo wa kialgorithmu uitwao Bayesian Learning Model unatumia kujifunza

kanuni za uondoaji utata wa kikompyuta kioto. Baadaye alitumia mbinu za Ramani ya Kujizatiti yenyewe (Self Organizing Map) ambao pia ni mfumo a kiagorithmu katika kuweka nomino za Kiswahili katika makundi yenye sifa sawa za kisemantiki.

Katika kujenga muonekano au mandhari ya kisemantiki ya nomino za Kiswahili aliangalia jinsi ya kuibuka na vigezo vya kusaidia kutofautisha maana na fasiri mbalimbali za maneno ya lugha pokezi (nomino) kutumia tafsiri-mashine kutoka lugha chanzi kwa mfano Kiingereza hadi Kiswahili. Kwa mfano nomino yenye utata wa maana kama "kaa" inaweza kuwa ni charcoal au crab katika tafsiri yake kwa Kiingereza ila sasa kigezo cha *WSD* cha kuonyesha sifa ya kisemantiki ya *ANIMATE* yaani iliyo hai kama kitengo cha kisemantiki (Ng'ang'a) itabainisha ni 'kaa' gani ilikuwa inarejelewa kisemantiki.

Sehemu hii itatoa mifano kadhaa kuonyesha jinsi uwekaji makundi ya kisemantiki unavyosaidia kuchunguza utatuzi wa utata wa maana kutumia mfumo unaojifunza lugha kikompyuta kwa misingi ya tafsiri za Kiswahili-Kiingereza.

MFANO 28

Class	Nouns	Occurrences
Abstract	93	70553
Animal	37	1963
Artifact	49	10238
Container	12	605
Dress	13	1622
Food	42	3718
Human	32	45993
Institution	32	31587
Location	46	23758
Money	20	6173
Plant	34	334
Substance	38	3923
Time	30	25349
Unit	29	1580
Vehicle	12	2906

Data asilia kutoka kwa Ng'ang'a 2005:57

Kikundi cha Kisemantiki	Idadi ya nomino	Mara ya utokeaji wa nomino katika kongoo
Kidhahania	93	70553
Wanyama	37	1963

Kifaa	49	10238
Chombo	12	605
Mavazi	13	1622
Chakula	42	3718
Binadamu	32	45993
Taasisi	32	31587
Mahali	46	23758
Pesa	20	6173
Mimea	34	334
Vitu	38	3923
Wakati	30	25349
Sehemu	29	1580
Magari	12	2906

Tafsiri yetu: Kutoka kwa Ng'ang'a 2005: 57

Katika mfano huu tunapata mandhari ya kijumla ya nomino za Kiswahili kisemantiki katika migao yake ikionyesha aina nomino katika makundi yake, idadi ya kila mgao wa nomino hizo na mara ngapi nomino hizo zimetokea katika kongoo husika.

MFANO 29

Tag I	Tag II	Word	Reading I	Reading II
Human	Artifact	<i>Gumegume</i>	Worthless person	Flint gun
		<i>Kiongozi</i>	Leader, guide	Manual, handbook
		<i>Mlezi</i>	Guardian, custodian	Cot
	Location	<i>Pepo</i>	Demon, spirit	Paradise
		<i>Bucha</i>	Butcher	Butchery
	Animal	<i>Kirukanjia</i>	Prostitute	Nightjar
		<i>Sungusungu</i>	Homeguard, vigilante	Black ant
<i>Mkunga</i>		Midwife	Eel	
Food	<i>Jimi</i>	Genie, wicked person	Gin	
	<i>Nyanya</i>	Grandmother	Tomato	
Plant	<i>Mtini</i>	Clown, buffoon	Fig tree	
Time	<i>Juma</i>	Name of person	Week	
Abstract	<i>Kichaa</i>	Lunatic	Lunacy	
	<i>Mwanga</i>	Wizard	Light	
	<i>Nyange</i>	Fool, moron	Noise	

Data asilia kutoka kwa Ng'ang'a 2005:60

Mgao wa 1	Mgao wa 1	Neno	Maana ya kwanza	Maana ya Pili
Binadamu	Kitu	<i>Gumegume</i>	Worthless person	Flint gun
	Mahali	<i>Pepo</i> <i>Bucha</i>	Demon, spirit Butcher	Paradise Butchery
	Wanyama	<i>Kirukanjia</i>	Prostitute	Nightjar

		<i>Sungusungu Mkunga</i>	Homeguard, vigilante Midwife	Black ant Eel
	Vyakula	<i>Jini Nyanya</i>	Genie wicked person Grandmother	Gin Tomato
	Mimea	<i>Mtini</i>	Clown buffoon	Fig tree
	Saa	<i>Juma</i>	Name of person	Week
	Kidhahania	<i>Kichaa Mwanga Nyange</i>	Lunatic Wizard Fool, Moron	Lunacy Light Noise

Tafsiri yetu: Kutoka kwa Ng'ang'a 2005:60

Huu ni mfano wa jinsi nomino chache katika kitengo Binadamu/Human zinavyoonyesha utata wa maana

MFANO 30

Uchanganuzi wa utata wa kileksika katika tafsiri za nomino za Kiswahili kwa Kingereza

Tag I	Tag II	Word	Reading I	Reading II
Animal	Vehicle	<i>Ndege</i>	Bird	Aeroplane
	Artifact	<i>Simu</i>	Sardine, sprat	Telephone
	Body	<i>Koo</i>	Hen, breeding animal	Throat, gullet
	Food	<i>Tembe</i>	Hen	Tablet
		<i>Kima</i>	Black monkey	Minced meat
	Abstract	<i>Swala</i>	Gazelle	Prayer
		<i>Goma</i>	Hard-skinned fish	Stick dance
		<i>Kima</i>	Black monkey	Price, value, rate
	Money	<i>Mbango</i>	Warthog	Money
	Dress	<i>Buibui</i>	Spider	Purdah, veil
	Time	<i>Mkasha</i>	Sparrow	Eve, vigil
	Container	<i>Chungu</i>	Black ants	Pot
	Location	<i>Kovongo</i>	Stork, crane	Gulley, ravine
<i>Barabara</i>		Crowned hornbill	Highway, road, street	
<i>Paa</i>		Gazelle	Roof	

Table 3.12: Ambiguity Group: Animal

Data asilia kutoka kwa Ng'ang'a 2005:60

Mgao wa 1	Mgao wa 1	Neno	Maana ya kwanza	Maana ya Pili
Wanyama	Gari	<i>Ndege</i>	Bird	Aeroplane
	Kitu	<i>Simu</i>	Sardine, sprat	Telephone
	Mwili	<i>Koo</i>	Hen, breeding animal	Throat, gullet
	Chakula	<i>Tembe</i>	Hen	Tablet

	Kidhahania	<i>Swala</i>	Gazelle	Prayer
	Pesa	Kima Mbango	Black monkey Warthog	Price, value, rate Money
	Wakati	<i>Mkeshia</i>	Sparrow Eve	vigil
	Nguo	<i>Buibui</i>	Spider	Purdah, veil
	Chombo	<i>Chungu</i>	Black ants	Pot
	Mahali	<i>Korongu Barabara Paa</i>	Stork, crane Gulley Crowned hornbill Gazelle	Ravine Highway Roof

Huu mfano nao unaonyesha jinsi nomino kadhaa katika kitengo Wanyama/Animal zinavyoonyesha utata wa maana

MFANO 31

Tag I	Tag II	Word	Reading I	Reading II
Location	Artifact	<i>Mkoa Komeo Mto</i>	Province, region Creek, inlet River	Metal bar Bolt, latch Pillow
	Food	<i>Kiwanda Tembe</i>	Factory House	Omelette Tablet
	Time	<i>Mwezi Magharibi</i>	Moon West	Month Sunset
	Body	<i>Ziwa</i>	Lake	Breast
	Plant	<i>Kambi Ua</i>	Camp Yard	Cambium Flower
	Vehicle	<i>Dau</i>	Pool	Dhow, sailboat
	Money	<i>Pango</i>	Cave	Rent
	Abstract	<i>Njia Kitende</i>	Road Residence, abode	Method, means Elephantiasis

Data asilia kutoka kwa Ng'ang'a 2005:61

Huu ni mfano unaoonyesha jinsi nomino kadhaa katika kitengo Nomino za Mahali/Location huwa zinavyoonyesha utata wa maana.

Mgao wa 1	Mgao wa 1	Neno	Maana ya kwanza	Maana ya Pili
Mahali	Kitu	<i>Mkoa Komeo Mto</i>	Province, Region Creek, Inlet River	Aeroplane Bolt, Latch Pillow
	Chakula	<i>Kiwanda Tembe</i>	Factory House	Omelette Tablet
	Wakati	<i>Mwezi Magharibi</i>	Moon West	Month Sunset
	Mwili	<i>Ziwa</i>	Lake	Breast
	Mimea	<i>Kambi Ua</i>	Camp Yard	Cambium Flower
	Gari	<i>Dau</i>	Pool	Dhow, Sailboat
	Pesa	<i>Pango</i>	Cave	Rent
	Kidhahania	<i>Njia Kitende</i>	Road Residence, abode	Method, Means Elephantiasis

MFANO 32

Tag I	Tag II	Word	Reading I	Reading II
Abstract	Plant	<i>Chacha</i> <i>Dege</i> <i>Mti</i>	Ballroom dance Convulsions, stomach pain Scrofula, gangrene	Grass Fern Tree
	Artifact	<i>Usukani</i> <i>Breki</i> <i>Useja</i> <i>Kifungo</i> <i>Mwiko</i>	Leadership Break, recess Celibacy, bachelorhood Detention Taboo, totem	Rudder, steering wheel Brake Collar Button Wooden spoon
	Time	<i>Magharibi</i> <i>Alasiri</i> <i>Alfajiri</i>	(sunset) Prayer (afternoon) Prayer (morning) Prayer	Sunset Afternoon Morning
	Food	<i>Zambarau</i> <i>Bia</i>	Purple Cooperation, agreement	Damson plum Beer
	Substance	<i>Madadi</i> <i>Ambo</i>	Assistance, support Disease	Opium Gum, glue
	Dress	<i>Doria</i> <i>Dibaji</i>	Security patrol Preface, preamble	Organdie, muslin Woollen/silk material
	Container	<i>Tusi</i>	Insult, abusive remark	Coffin, bier
	Body	<i>Sini</i>	Complexion, shape	Gum (of teeth)
	Vehicle	<i>Jipu</i>	Boil, abscess	Jeep

Data asilia kutoka kwa Ng'ang'a 2005:61

Kitengo cha Nomino za Kidhahania/Abstract na hali za utata wa maana.

Data asilia kutoka kwa Ng'ang'a 2005:61

Mgao wa 1	Mgao wa 1	Neno	Maana ya kwanza	Maana ya Pili
Kidhahania	Mimea	<i>Chacha</i> <i>Dege</i> <i>Mti</i>	Ballroom dance Convulsions, stomach pain Scrofula, gangrene	Grass Fern Tree
	Kifaa	<i>Usukani</i> <i>Breki</i> , <i>Useja</i> , <i>Kifungo</i> <i>Mwiko</i> ,	Leadership Break, recess Celibacy bachelorhood Detention Taboo, totem	Rudder, steering wheel Brake Collar Button Wooden spoon
	Wakati	<i>Magharibi</i> <i>Alasiri</i> <i>Alfajiri</i>	(sunset) Prayer (afternoon) Prayer (morning) Prayer	Sunset Afternoon Morning
	Vyakula	<i>Zambarau</i> <i>Bia</i>	Purple Cooperation or agreement	Damson Plum Beer
	Kitu	<i>Madadi</i> <i>Ambo</i>	Assistance, support Disease	Opium Gum, glue
	Mavazi	<i>Doria</i> <i>Dibaji</i>	Security Control Preface, Preamble	Organdie, muslim Woollen/silk material
	chombo			
	Mwili			

Huu ni mfano unaoonyesha jinsi nomino kadhaa katika kitengo Nomino za kidhahania/ Abstract huwa zinavyoonyesha utata wa maana.

MFANO 33

Noun	Classes (senses)	Auto	Bias	Test	Bias
		C1	C2	C1	C2
<i>juma</i>	HUMAN-TIME	59.6	40.4	80.3	19.7
<i>mkunga</i>	HUMAN-ANIMAL	95.6	4.4	97.1	2.9
<i>kiongozi</i>	HUMAN-ARTIFACT	79.7	20.3	97.8	2.2
<i>kirukanjia</i>	HUMAN-ANIMAL	95.6	4.4	33.3	66.7
<i>nyanya</i>	HUMAN-FOOD	90.1	9.9	6.3	93.7
<i>korongo</i>	ANIMAL-LOCATION	7.0	93.0	68.4	31.6
<i>ndege</i>	ANIMAL-VEHICLE	35.1	64.9	72.6	27.4
<i>buibui</i>	ANIMAL-DRESS	52.0	48.0	43.3	56.7
<i>tembe</i>	LOCATION-FOOD	84.7	15.3	83.3	16.7
<i>ua</i>	LOCATION-PLANT	98.9	1.1	63.8	36.2
<i>mwezi</i>	LOCATION-TIME	47.3	52.7	19.2	70.8
<i>pango</i>	LOCATION-MONEY	77.9	22.1	70.3	29.7
<i>sindano</i>	ABSTRACT-ARTIFACT	87.3	12.7	28.2	71.8
<i>saa</i>	ARTIFACT-TIME	27.3	72.7	5.6	94.4
<i>bakora</i>	ABSTRACT-ARTIFACT	87.3	12.7	40.5	59.5
<i>usukani</i>	ABSTRACT-TIME	87.3	12.7	47.1	52.9

Data asilia kutoka kwa Ng'ang'a 2005:61

Data iliyokusanywa kutoka kwa kongoo ya kujifunza lugha ikionyesha ujaribishaji wa nomino mbalimbali na viwango vya mutokeaji wa utata wa maana.

MFANO 34

Noun	Type	Auto		None		Test		Feature Set
		MFS	ACC.	MFS	ACC.	MFS	ACC.	
<i>juma</i>	Homo	59.6	84.0	50	71.9	80.3	83.8	W+L+M
<i>mkunga</i>	Homo	95.6	97.1	50	94.1	97.1	97.1	M
<i>kiongozi</i>	Meta	79.7	97.8	50	97.8	97.8	97.8	M,L+M
<i>kirukanjia</i>	Meto	95.6	33.3	50	66.7	66.7	66.7	W,L,M,L+M
<i>nyanya</i>	Homo	90.1	12.5	50	93.8	93.7	93.8	M
<i>korongo</i>	Homo	93.0	57.9	50	84.2	68.4	84.2	M
<i>ndege</i>	Meta	64.9	81.2	50	80.1	72.6	82.3	M
<i>buibui</i>	Homo	52.0	66.7	50	70.0	56.7	66.7	L+M
<i>tembe</i>	Homo	84.7	100	50	100	83.3	83.3	L+M
<i>ua</i>	Homo	98.9	63.8	50	63.8	63.8	74.5	L+M
<i>mwezi</i>	Meta	52.7	85.9	50	88.5	70.8	82.1	L
<i>pango</i>	Homo	77.9	78.2	50	75.2	70.3	79.2	L+M
<i>sindano</i>	Meta	87.3	28.2	50	71.8	71.8	76.9	L
<i>saa</i>	Meto	72.7	91.6	50	86.9	94.4	94.4	L
<i>bakora</i>	Meto	87.3	40.5	50	54.1	59.5	59.5	W
<i>usukani</i>	Meta	87.3	70.6	50	64.7	52.9	64.7	M
AVERAGE		80.0	68.1	50.0	79.0	75.0	80.4	

Data asilia kutoka kwa Ng'ang'a 2005:96

Baada ya migao hii nah ii ya nomino katika kategoria mbalimbali na kupitishia data hii katika programu kadhaa tulizozitaja, ilikuwa dhahiri kuwa bado kulikuwa na kazi ya kufanywa ila utata wote uliokuwa wazi na hata kujaribiwa kiasilimia mathalan katika mfano wa 33 na 34. Tukimulika neno ‘juma’ kwa mfano katika mfano wa 33 lipo kwenye makundi mawili ya kisemantiki- kundi la wakati- TIME- na kundi la binadamu – HUMAN.

Ili kuondoa utata huu wa kisemantiki, Ng’ang’a 2005:96 ametumia programu ya SWATWOL ya SALAMA kupata muktadha wa neno ‘juma’ ili kulitalii kisemantiki. Mkutadha wenyewe upo katika Mfano wa 35 uliotokana na kongoo ya SALAMA matini ya sehemu ya Biblia kutoka kwa Agano jipya. Neno ‘juma’ limeainishwa kama ‘wiki’ na pia kama ‘jina’ la mtu.

MFANO 35

naye alipofufuka and when he resurrected	alfajiri siku ya kwanza ya in the morning of the first day of	juma , week/name ,	alimtokea kwanza (4.31) he appeared first to
mariammu magdalena mary magdalene	ambaye kwamba alimtoa pepo saba from whom he cast out seven demons		

Data asilia kutoka kwa Ng’ang’a 2005:99

MFANO 36

"<kwanza>" "kwanza" NUM NUM-INFL ORD { first }
"<ya>" "ya" GEN-CON 9/10-SG
"<juma>" "juma" N 5a/6-SG { week } AR
"<>" "," COMMA
"<alimtokea>" "tokea" V 1/2-SG3-SP VFIN PAST 1/2-SG3-OBJ OBJ { put out , remove , publish , produce , generate , offer to someone , subtract , reduce } SVO EXT: STAT APPL :EXT

Mchoro:2. Uchanganuzi wa SWATWOL kuhusu jinsi ‘juma’ linavyoondoshewa utata wa kisemantiki

Mfano huu wa 36, imetumia uainishaji wa SWATWOL na kuondoa utata wa kisemantiki wa neno 'juma' na kubainisha kuwa katika muktadha huu ni 'wiki'. Kinachoonekana wazi ni kuwa muktadha bado unasaidia katika uchunguzi wa viwango vyote vya lugha kutumia teknolojia ya lugha hasa pale matini inayochunguzwa au kutafitiwa inatokana na kongoo iliyoainishwa kama HCS au kutumia programu zenye msamiati wake kama SWATWOL. Tafsihi imejitokeza kama kipengele muhimu kwa taaluma ya semantiki kwa kuwa inasaidia ubainishaji wa maana lengwa na utatuzi wa utata unaotokea katika lugha chanzi.

4.3.3 Kupata viwango vya ufasaha katika utatuzi wa utata wa kisemantiki.

Ni dhahiri kuwa mkabala huu wa mbinu mseto za kiteknolojia ya lugha alizitumia Ng'ang'a 2005 katika utafiti wa kisemantiki wa Kiswahili una mchango katika taaluma nzima ya teknolojia ya lugha ya Kiswahili. Ameweka msingi imara katika uchanganuzi wa kisemantiki wa nomino za Kiswahili na ujenzi wa vigezo vya kuainisha maneno hasa nomino za Kiswahili kwa msingi wa kategoria kwa kutumia SOM- Ramani ya Kujizatiti yenye inayowakilisha mandhari halisi ya nomino za Kiswahili yanayoonyesha utokeaji wake na mfafano wa sifa za kisemantiki katika nomino hizo(Ng'ang'a 2005).

Pia data na uafafanuzi wa utafiti huu umeonyesha jinsi ya kuunda kiainishi cha kisemantiki au mfumo wa kikompyuta wa kuainisha nomino za Kiswahili kwa msingi wa kujenga mandhari ya kisemantiki yaliyojengwa katika SOM, mbinu ambayo inaweza kufanyia utafiti msamiati wa Kiswahili pakubwa na pia kusaidia uchunguzi wa matawi mengine ya kiisimu kama sintaksia na mofolojia. Kongoo ya HCS ya SALAMA na kamusi ya TUKI vimesaidia katika mchakato huu wa utafiti wa uondoaji utata wa maana ya maneno ya Kiswahili na pia uchunguzi wa miktadha ya utata uliojitokeza ulifanywa kwa matumizi ya programu ya SWATWOL ambayo tumeona uwezo wake tayari.

4.4. Teknolojia ya lugha katika utafiti wa Sintaksia ya Kiswahili

Kwa ujibu wa TUKI, 1990 Sintaksia ni tawi la sarufi linaloshughulikia mpangilio na uhusiano wa vipashio katika sentensi. Kwa hivyo lazima maneno yapangwa kwa njia inayokubalika katika sentensi ndipo yaitwa lugha. TUKI, 1990:55 wanazungumzia dhana ya tungo-kisintaksi ambayo wanasema ni matokeo ya kupanga vipashio vya muundo kulingana na kanuni za kisarufi za lugha

hiyo. Lugha tunayozungumza kila siku au kuiandika iaitwa lugha kwa kuwa inafuatia utaratibu huu. Lugha hii ikiwekwa katika muundo wa matini za kusomeka kikompyuta na kuhifadhiwa katika seva za kompyuta, huitwa kongoo, na inaweza kuchunguzwa kutumia mbinu za teknolojia ya lugha.

Kama tulivyofafanua kwa kirefu, Kiswahili ni lugha inayotumia miundo changamano sana ya kimofolojia. Miundo hii huwa ni hazina kuu kwa uchunguzi wa sintaksia kwa kuwa masuala ya miundo ya maneno, uambishaji wake huathiri pakubwa yanavyopangwa katika sentensi (Hurskainen 2009:6) na yanavyopangwa katika sentensi huathiri maana yake kisemantiki. Kwa hivyo ni wazi kuwa viwango hivi vyote vya sarufi hushikana, kuathiriana na kuhusiana pakubwa. Tumeyaona mahusiano haya katika mifano yetu ya kueleza kuhusu teknolojia ya lugha na ilivyotumika katika mradi wa SALAMA.

Katika moja ya programu muhimu tulizoeleza ni SWAFDP ambayo ni akronimu ya Swahili Functional Dependency Parser ambacho ni kama ilivyoielezwa ni kichambuzi cha kikompyuta kilicho na uwezo mkubwa kuliko SWACGP- sarufi kikwazo na hufanya uchanganuzi wa kina wa kisintaksia na huweza hata hufafanua Kiswahili kwa kielelezo cha miti. Hurskainen2004:376 anafanua kuhusu uchambuzi wa kisintaksi wa kijuujuu na ule wa kina. Tutatoa mifano ya uchunguzi wa kisintaksi kwa kina na kijuu juu pia inapohitaji kuweka msingi wa wazo lengwa.

Kama tulivyofafanua katika sura ya pili, SALAMA iliweza kushirikisha kipengele cha SWAFDP au Kichambuzi cha Sarufi Amilifu Tegemezi ya Kiswahili kama ilivyokuwa imeundwa na wenzake Järvinen and Tapanainen (1997) kuchunguza sintaksia ya Kiswahili kwa kutumia msamiati ulio kwenye SWATWOL na matini za kongoo ya HCS kwa jumla (Hurskainen 2004:376). Mkabala huu ulifaa kuchunguza sintaksia kwa misingi ya sifa hizi nane;

- i) Kipashio cha kimsingi cha kisintaksia huwa si neno ila ni kiini chake
- ii) Kila kipashio cha kisintaksia kina kichwa kimoja.
- iii) Matokeo ya mchakato huu huwa ni kielelezo cha miti
- iv) Utegemezi wa kiuamilifu huwa unaelezewa kupitia kwa vishirikishi
- v) Vishirikishi hivi vinaweza kukutania mahali kwenye tungo la kisarufi
- vi) Vivumishi havielezei dhamira
- vii) Sarufi huwa si zalishi kwa hivyo programuya SWAFGP hupokea sentensi ya aina yoyote na kuichanganua na kupata hali yake ya kisintaksia hata kama sentensi si sahihi kisarufi.
- viii) Ufafanuzi wa kiutegemezi huwa ni wa kiwango kimoja (monostratal) cha kisintaksia yaani maelezo ya kimsingi bila ugeuzajimaumbo.

Uzuri wa kichambuzi hiki cha sarufi tegemezi ni kuwa miundo ya kisintaksi huelezewa kwa

udhahiri kabisa kiasi cha hata kuweza kutambua maneno katika sentensi ambayo ina tungo-funge hata yakiwa hayafuatani katika sentensi. Uwezo huu husaidia utambuzi wa kila aina ya neno au makundi ya maneno katika sentensi yoyote ya Kiswahili.

MFANO 37

Hebu tuangalie sentensi hii ...*Mtoto analima shamba*- jinsi ilivyochanganuliwa kisintaksia kwanza kijuujuu kasha baadaye kwa kina katika MFANO 38.

```
... "<*mtoto>"
"mtoto" N CAP 1/2-SG { child , young person , juvenile } @SUBJ>
"<analima>"
"lima" V 1/2-SG3-SP VFIN { he/she } PR:na z [lima] { cultivate , dig , farm , plough
, till , hoe } SVO @FMAINVtr+OBJ>
"<shamba>"
"shamba" N 5a/6-SG { farm , field , plot for cultivation , plantation , estate ,
countryside } @OBJ<
"<.$>"
"." { .$ } ..."
```

Data asilia kutoka Hurskainen 2004:375-376

MFANO 38

```
main#1
*mtoto
"mtoto" CAP N 1/2-SG { child , young person , juvenile } @NH >2
analima
"lima" V 1/2-SG3-SP VFIN PR:na z { dig , farm , cultivate , plough , till , hoe } SVO
@MAIN obj#3 subj#2 >1
shamba
"shamba" N 5a/6-SG { farm , field , plot for cultivation , plantation , estate , country
side } @NH >3
.$
```

Data asilia kutoka Hurskainen 2004:375-376

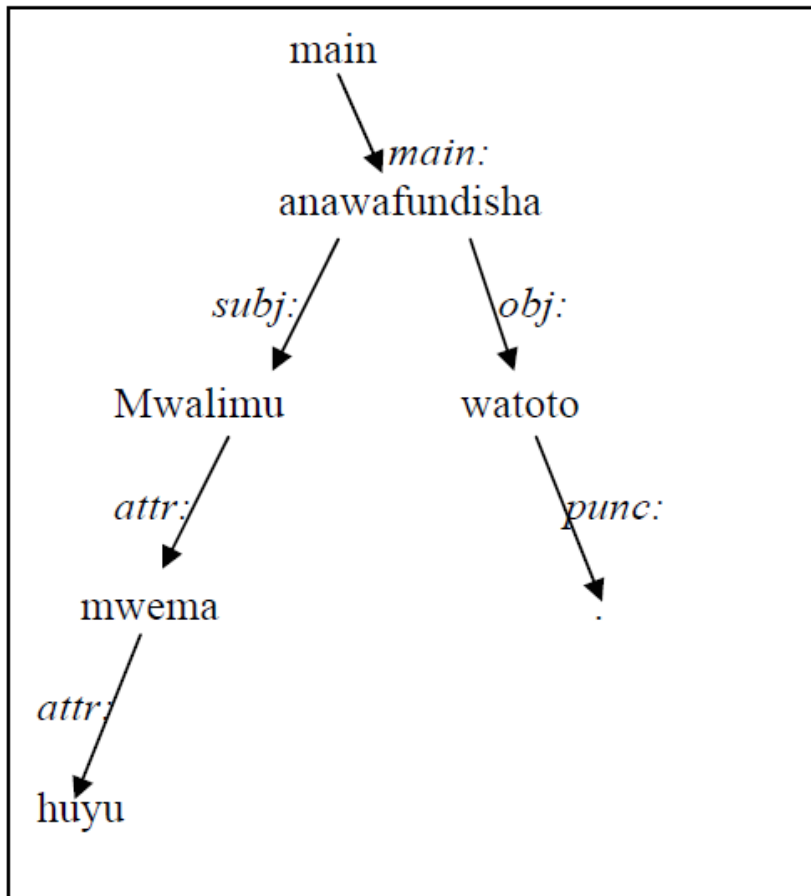
Kisha tuangalie mfano tofauti katika MFANO 39- *Mwalimu huyu mwema anawafundisha watoto.*

MFANO 39

```
main#1
*mwalimu
"mwalimu" N CAP 1/2-SG { a/the } { teacher } @NH attr#2 >3
huyu
"huyu" PRON DEM :hV 1/2-SG { this } @POSTMOD attr#4 >2
mwema
"ema" ADJ A-INFL 1/2-SG { good } @POSTMOD >4
anawafundisha
"fundisha" V 1/2-SG3-SP VFIN { he/she } PR:na 1/2-PL3-OBJ OBJ { them } z [funda] { teach ,
instruct , inculcate , indoctrinate } SVO EXT: SVO-C CAUS :EXT @MAIN obj#5
subj#3 >1
watoto
"mtoto" N 1/2-PL { the } { child , young person , juvenile } @NH >5
.$
"." { .$ }
```

Data asilia kutoka Hurskainen 2004:375-376

MFANO 40a



Data asilia kutoka Hurskainen 2004:375-376

MFANO 40b

Uainishaji wa kisintaksia wa sentensi hii *Nyakati hizi kila mtafiti analazimika kupambana na wafadhili ili apate mbinu za kufanya kazi ya utafiti.*

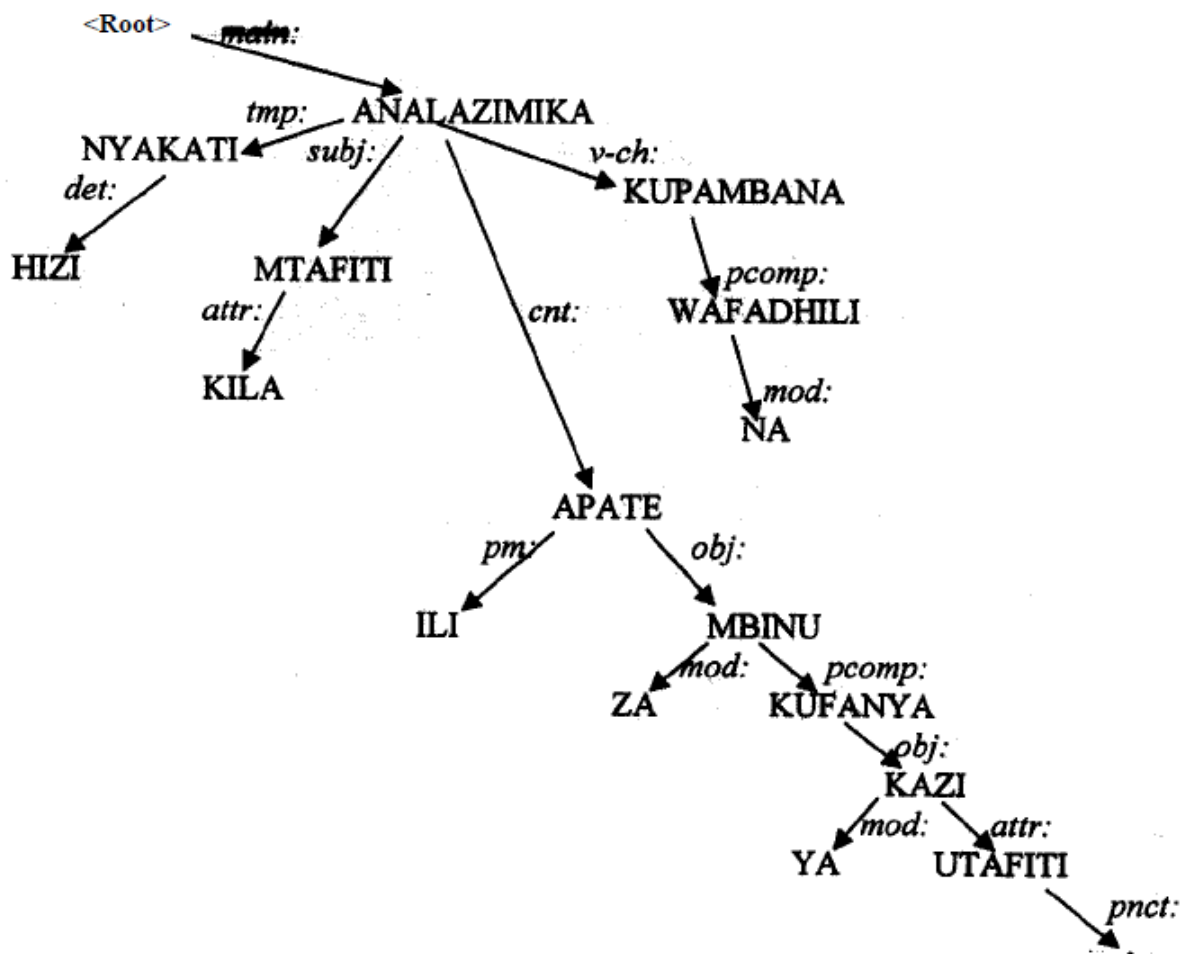
Ufafanuzi wa kisintaksia kwa njia kusomeka kikompyuta

- 1 *nyakati wakati tmp:>5 @ TIME 11/10-PLNAR' time '
- 2 hizi hizi det:>1 @ <ND PRON:DEM hV 9/10-PL ' this '
- 3 kila kila det:>4 @ AD-A> AR A-UNINFL ' every '
- 4 mtafiti mtafiti subj:>5 @ SUBJ ½-SG N AR ' researcher '
- 5 analazimika lazimika main:>0 @ FMAINV ½-SG3-SP VFIN:PRna V
AR SV ' be forced to ' STAT
- 6 kupambana pambana mod:>5 @ -FMAINV-n INF V SV SVO '
fight with ' REC
- 7 na na cc:>8 @ CC CC ' with '
- 8 wafadhili mfadhili obj:>6 @ I-OBJ ½-PL N AR ' sponsor '
- 9 ili ili pm:>10 @ CS AR**CLB CONJ ' in order to '
- 10 apate pata cnt:>6 @ FMAINVtr> SBJN VFIN ½-SG3-SP V
SV SVO ' get '
- 11 mbinu mbinu obj:>10 @ OBJ 9/10-NI-PL N ' means '
- 12 za za mod:>11 @ <NOM 9/10-PL GEN-CON
- 13 kufanya fanya pcomp:>11 @ -FMAINV-n INF V SV SVO ' do '

14 kazi kazi obj:>13 @OBJ 9/10-0-SG N ' work '
 15 ya ya **:>16 @ <NOM/10-SG GEN-CON
 16 utafiti utafiti attr:>14 @ <P 11-SG N AR HC ' research '.\$

Data asilia kutoka Hurskainen 1999:154

Ufafanuzi wa kisintaksia wa miti kwa kwa wasilisho la muonekano wake



Data asilia kutoka Hurskainen 1999:155

4.4.1 Uchanganuzi wa data

MFANO 37- 40 ni muhimu inadhihirisha njia za kuchanganua tungo za Kiswahili kisintaksia kwa utambulisho wa aina mbili wa vipengele mbalimbali vya kisintaksia unavyofanyika kwa maneno tofauti katika sentensi wa kwanza ukiwa ni wa kiambishi @ kinachosawiri sifa za kisintaksia za neno moja moja bila urejelehi wa sentensi yote Hurskainen (2004:377). Pili kuna utambulisho wa kuonyesha viungo kati ya maneno kwa '>nambari' na pia vichwa "#nambari".

Kwa mfano *subj#2* huumanisha kuwa kipengele hicho ni kichwa cha kiima ambacho kipo katika nafasi ya pili. Nacho kialamishi >2 kinamaanisha kiungo kinachounganishwa na kichwa chenye nambari inayolingana na hiyo na pia *obj#3* inaashiria kichwa cha yambwa na yambwa yenyewe inalamishwa na >3 ambacho ni kiungo baina ya hizo mbili. Kitenzi cha *analima* katika Mfano 38 ndicho kichwa cha Kirai Kitenzi au neno kuu hapo na inatokea juu ya muundo huo (sentensi) na hivyo kuwa kiungo kikuu baina ya kiima na yambwa >1 kama ilivyo katika mfano wa 38. (Hurskainen 2004:378).

Katika sentensi *Mwalimu huyu mwema anawafundisha watoto* katika mfano wa 39 kwa hivyo tunapata kuwa baadhi ya vibainisha vya nafasi ya neon katika sentensi au uamilifu wa neon hilo ndio muhimu katika uhusishwaji na vibainishi vya kimofolojia ambavyo vinatenda majukumu ya kisintaksia. Kama neno *mwalimu* lina alama ya kichwa kama *attr#2* ambayo imeungwa na *huyu* (>2) ambalo neno *huyu* tena ni ambacho ni kivumishi kionyeshi ni kichwa (*attr#4*) kwa kivumishi *mwema* na hivyo neno *mwema* si kichwa cha sehemu yoyote sentensini na ndiyo sababu hakina utambulisho au alama ya kichwa.

Kitenzi *anawafundisha* bila shaka ni kichwa cha kwa kiima (*subj#3*) na hivyo imeunganishwa nacho kwa (>3). Kitenzi kicho hicho ni kichwa pia cha yambwa (*obj#5*) na zimeungwa pamoja na (>5). Yambwa watoto si kichwa na hivyo haina kinachoitambulisha hivyo. Kielelezo cha miti katika mfano 40 ni njia tofauti ya kuwasilisha sentensi hii ya MFANO 39 na hakuna tofauti baina ya uchanganuzi katika mifano yote miwili Hurskainen (2004:378).

Kwa hivyo, kwa mujibu wa data hii, ni dhahiri kuwa matumizi ya programu anuwai katika kuchanganua lugha ya Kiswahili kisintaksia yameonyesha kuwa na udhahiri zaidi kuliko uchanganuzi wa kiungamuzi ambapo mtaalamu huainisha neno kwa kuliangalia na kulipa nafasi au uamilifu wake. Ithibati kuwa kila neno katika kongoo ni kipashio cha kikompyuta na limechanganuliwa kimofolojia, ni dhahiri kuwa data itakayoibuka inaaminika.

Pili, data inayoibuka katika uchanganuzi wa kisintaksia kutumia mbinu za kimofolojia kama SWATWOL, sarufi kikwazo yaani SWACG, kutumia Sarufi Amilifu Tegemezi SWAFDP na mbinu za kuondosha utata wa kisemantiki katika tungo za Kiswahili, yamedhihirisha sio tu katika mifano mine hii ila pia katika sura ya pili kuwa, matumizi ya uahadisi lugha au mbinu za kiteknolojia ya lugha, zinashughulikia lugha kwa njia kadhaa na kwa kina zaidi. Kwa mfano katika mfano wa 39 na 40, tunaweza kutambua, siyo tu nafasi za kisintaksi za baadhi ya maneno lakini pia yanapounganishwa, kama ni kichwa cha sentensi au kirai hicho, je, ni kivumishi kipi cha kiima ni

kichwa katika mpangilio huo na haya yote ni kwa njia ya hakika inayoweza kujaribishwa kisayansi bila kukosea mchakato uliopitia.

Sambamba na nadharia ya Msambao wa ugunduzi, uchakataji huu hutumia sekunde chache kupitisha sentensi hiyo kwa kompyuta. Pia mbinu za kikompyuta za kufanyia utafiti Kiswahili kisintaksia siyo hizi tu. Ikumbukwe pia kuwa katika utafiti wa semantiki tulisema kuwa zana kama SOM-Self Organizing Map au Ramani Inayojizatiti Yenyewe anayoitumia Ng'ang'a 2003 na 2005 inapanga viima na vitenzi na kueleza uhusiano wao, njia ya hakika pia ya kufanyia utafiti sintaksi ya lugha ya Kiswahili (Hurskainen 2005:122).

4.5 Teknolojia ya Lugha katika Uundaji na Utathmini wa Kamusi za Kiswahili

Katika sura ya pili, tulieleza kijuujuu jinsi mradi wa SALAMA ulivyotumia programu ya SALAMA-DC katika uundaji na utathmini kamusi za Kiswahili. Sehemu hii itaonyesha mifano na kutoa ufafanuzi wa jinsi SALAMA-DC ulivyo mfumo bainifu wa kuainisha lugha na jinsi unavyoweza kutumika katika kuchakata kwa haraka na kioto, hatua nyingi, za kiholela na za kuchosha zinazohusika katika uundaji wa kawaida wa makamusi ya Kiswahili.(Hurskainen, 2008:2).

Tutatoa na kueleza data za matumizi ya teknolojia ya lugha katika uundaji na utathmini wa kamusi ya Kiswahili kwa kutolea mifano mbalimbali programu ya SALAMA-DC iliyofaulu kafanya yafuatayo; kutoa kidahizo cha neno moja, kutoa kidahizo cha tungo-funge, kuibuka na mfumo ya kurejeleana na hata kuteua mifano katika miktadha mbalimbali, pamoja na uteuzi nasibu na uteuzi unaolingana na miktadha ya utokeaji wa mifano hiyo.

Kwa hivyo, katika kutimiza majukumu ya kuunda ‘...mfumo thabiti na bayana wa kuzalisha vidahizo vya kamusi kutoka kwa aina yoyote ya muundo wa maneno ya Kiswahili...’ kwa njia ya kikompyuta (Hurskainen, 2008:3), SALAMA-DC imeweza bila kujali aina za maneno au ukubwa wa kongoo-matini chanzi. Mfumo huu kwa hivyo hutoa vidahizo au maneno makuu ya kamusi inayoundwa kwa kutumia yale maneno tu yanayopatikana katika kongoo-matini pamoja na ²²habari za uchanganuzi wa maneno kiidadi (kuonyesha neno fulani limetokea mara ngapi). Hurskaninen anaonyesha kuwa maneno ambayo ni nomino maalum ambazo hazijazoeleka na mengine ambayo ni makosa ya kitahajia yanayotambulika na hutengwa ili yashughulikiwe baadaye na watu.

4.5.1 Uundaji wa vidahizo vya kamusi

Programu ya SALAMA-DC inaundaje kamusi? Inafanya hivyo kwa kutekeleza majukumu mawili ambayo ni kutoa vidahizo vya kikamusi vilivyo na uainishaji wa kiisimu ufaao na pili ni kutafuta mifano ya kimatumizi ya maneno hayo katika kongoo-matini husika (Hurskainen 2008). Majukumu haya mawili pia yatakuwa msingi wa kutathmini programu ya kuunda kamusi na kamusi zenyewe katika sehemu hii.

Mfumo wa SALAMA-DC, unatumia matini za Kiswahili ambazo ni kongoo husika inayosomeka kikompyuta yaani (Hurskainen 2008) kumaanisha; zimechanganyuliwa kimofolojia, zikaondolewa utata wa kimofolojia na wa kisemantiki, kutenga na kufafanua tungo-funge kama nahau na misemo au maneno ya makundi, kutoa maneno na faharasa na maana zake kwa lugha ya Kiingereza na pia kufanya uchambuzi wa kisintaksia. Kwa hivyo mfumo huu huunda vidahizo kutokana na maneno tu ambayo ambayo yamo kwenye kongoo pamoja na habari za idadi za utokeaji wa maneno hayo

4.5.1.1 Uundaji wa vidahizo vya neno moja

Programu ya SALAMA-DC ina uwezo wa kuunda aina mbili za vidahizo vya neno moja vikiwa ni kidahizo cha mzizi wa kitenzi na pia vidahizo vya neno ambalo ni kitenzi kilichonyambuliwa

Kamusi nyingi za Kiswahili kwa kawaida huwa na vidahizo vya mzizi wa kitenzi tu bila kuongeza kila hali ya mnyambuliko wako. SALAMA-DC inashirikisha minyambuliko hii kama vidahizo baada ya kuvipitishia kwa mchakao wa kuvifafanua kileksika na kwa sababu ya muktadha wake vikuwa na maana mahsusii katika kamusi kwa mujibu ya mifano yake katika kongoo-matini.

MFANO 41

{abudia} headword as a lexical entry
(abudu) base form of verb

Data asilia kutoka kwa Hurskainen, 2008:6

(abudu) - shina la kitenzi husika
{abudia} – kidahizo katika kamusi

Data asilia kutoka kwa Hurskainen, 2008:6

.....
²²...orodha za uchanganuzi wa maneno kiidadi (frequency lists) yaani kuonyesha neno fulani limetokea mara ngapi katika matini... Hurskainen, 2004:39 ametoa mfano wa uchanganuzi huu anaouita 324120 na CC { and } 152569 kwa PREP { at, to, for },135749 wa V { be },135671 katika PREP { in, at },122973 sema V { say, speak, scold, speak against, advise, counsel, backbite, badmouth } ,96886 la ADV { no, not } AR ,63017 na PREP { with } 59806 nchi N 9/10 { country, land, ground } 57447 na AG-PART { by } 46636 mwaka N 3/4 { year }, 44812 mtu N 1/2 { human being, person, individual } Hii inamaanisha kuwa katika Kongoo-matini hiyo ya SALAMA kwa mfano neno –mtu limetokea mara 44,812 na ni nomino ya ngeli ya kwanza lenye tafsiri ya man, person na kadhalika.

4.5.1.2 Uundaji wa vidahizo vya tungo-funge

Tungo-funge kama tulivyoelezea mwanzoni ni maneno ambayo hutokea pamoja yakiwa yanarejelea dhana moja kwa kuwa kama vile nahau, misemo, methali na kadhalika na kwa hivyo maana ya tungo kama hizo haiwezi kafahamika kwa kueleza vijenzi vyake kimojakimoja. Kufikia 2008, SAMALA ilikuwa imetenga na kufafanua tungo-funge zenye nahau 2,127, methali 2,173, na aina nyingine tofauti za tungo-funge 6,338 (Hurskainen 2008:5) za Kiswahili kazi anayosema haikuwa rahisi kwa kuwa lugha hii inaweza kuwa na kitenzi kilichonyambuliwa katika nahau fulani. Alitumia uondoaji wa utata wa kimofolojia katika kuzitenga tungo-funge hizo ambazo zilikuwa ni pamoja na miundo mbalimbali ya nomino na vivumishi tofauti.

Methali pia hutengwa na kuwekwa katika sehemu tofauti ya kamusi. Ajabu ni kuwa makundi yote haya ya maneno huchotwa kutoka kwa kongoo na programu ya SALAMA-DC na kuyatolea mifano ya maana kwa miktadha ya matumizi yake katika lugha kwa mujibu wa kongoo iliyoshughulikiwa (Hurskainen 2008:5)

SALAMA-DC huyatambua je maneno haya ambayo ni tungo-funge? Kwa kutumia mifano aliyotoa Hurskainen 2008 katika Ripoti Nambari 2 ya 2008, tutafafanua jinsi SALAMA-DC hutambua neno tangulizi la tungo-funge na kutumia urejeleano katika kufanya kila kundi hilo kwa kidahizo.

MFANO 42

"... [Afyā] see [bwana_afya] [enye_afya] [enye_nguvu_na_afya]
[akili] see [enye_akili] [fanya_akili] [rukwa_na_akili] [teka_akili] [tia_akili]..."
Data asilia kutoka kwa Hurskainen, 2008
Ripoti Nambari 2 ya 2008:6

MFANO 43

'...[afya] see [bwana_afya] N 9/6 HUM MW-N { health officer } 10
[afya] see [enye_afya] ADJ MW> { bonny } 17
[afya] see [enye_nguvu_na_afya] ADJ MW>>> { hale } 1
[akili] see [enye_akili] ADJ MW> { clever, cute } 32
[akili] see [fanya_akili] V IDIOM-V { use thinking } 1
[akili] see [rukwa_na_akili] V PASS IDIOM-V { lose one's mind, be mentally ill } 2
[akili] see [teka_akili] V IDIOM-V { control completely } 1
[akili] see [tia_akili] V IDIOM-V { take note of } 1.."
Data asilia kutoka kwa Hurskainen, 2008
Ripoti Nambari 2 ya 2008:6

4.5.1.3 Uundaji wa vidahizo vya kipolisemia

Polisemia kwa mujibu wa TUKI 1990:17 ni hali ya neon au kundi la maneno lenye maana zaidi ya moja. Kwa hivyo kipashio cha kipolisemia huwa na zaidi ya fasiri moja ya kisemantiki na asemavyo Hurskainen 2008 ni kuwa tofauti za maana ya kipolisemia hudhihirika vyema kwa matumizi ya mifano inayofaa ya kila neno la kipolisemia kama inavyopatikana katika kongoo. Neno la kipolisemia linaweza kuwekwa katika kamusi kama kidahizo cha kawaida kama ifuatavyo;

MFANO 44

... {fedha} N 9/10-SG { 1 money, 2 silver, 3 brass, 4 coin, 5 capital, 6 finances }

Data asilia kutoka kwa Hurskainen, 2008

Ripoti Nambari 2 ya 2008:6

Kwa mujibu wa Hurskainen 2008, ili tupate mifano mwafaka katika matumizi itabidi kuondoa utata wa kisemantiki katika kila neon linalotokea na maana tofauti kama ilivyo kwenye mfano huu;

MFANO 45

‘... {fedha1} N 9/10-SG {money}

{fedha2} N 9/10-SG {silver}

{fedha3} N 9/10-SG {brass}

{fedha4} N 9/10-SG {coin}

{fedha5} N 9/10-SG {capital}

{fedha6} N 9/10-SG {finances}...’

Data asilia kutoka kwa Hurskainen, 2008

Ripoti Nambari 2 ya 2008:6

Programu ya SALAMA-DC kwa sasa inaweza kushughulikia hili na kiwango kikubwa cha polisemi zinaunganishwa kikompyuta na mifano yake ila mfumo huu haujakamilika kabisa na unazidi kuboreshwa ikiwa ni pamoja na kuuwezesha kuonda utata wa kisemantiki ili upate kuwa ufasaha zaidi Hurskainen 2008.

4.5.1.4 Uundaji wa vidahizo vya kihomonimu

TUKI 1990 inatoa fasili ya homonimi kuwa ni leksimu mbili au zaidi katika lugha moja zenye kuandikwa sawa lakini zenye maana tofauti. Programu ya SALAMA-DC ina uwezo wa kutambua kila homonimi na kutoa habari zake za idadi ya utokeaji kwenye matini. Utambuzi huu hutokea katika mchakato wa uondoaji utata wa kisemantiki ambapo sheria mbalimbali za kiisimu hutumika kikompyuta katika kufafanua maana ya kila homonimi (Hurskainen 2008:6 na kuitolea mifano yake kama ilivyoonyeshwa hapa chini. Kwa kuwa Kiswahili ni lugha yenye mfumo wa ngeli za nomino zake na hononimi zake ni chache ila ni sharti kila moja iwekwe kwenye kamusi kama kidahizo.

Tazama kwa mfano jinsi neno 'funza' ambalo ni nomino na pia kitenzi linavyoelezwa baada kutokea katika Kongoo 1 mara moja kama nomino na mara kumi kama kitenzi (Hurskainen 2008:6)

MFANO 46

‘...{funza} N 9/10 {sand flea, larva, maggot, grub } HUM 1
{funza} V (funza) {educate, teach good manners} 10...’

Data asilia kutoka kwa Hurskainen, 2008

Ripoti Nambari 2 ya 2008:6

Katika muktadha huu pia tunaona jinsi SALAMA-DC inavyoweza kuunda kidahizo na kutoa mifano kutoka kwa homonimi mbali katika kongoo tofauti.

MFANO 47

‘...jibu} N 5/6 { answer, reply, response } AR

[jibu] <DWE> wanahoji wafuasi wake wanaohakikisha jibu [jibu] litakuwa kali (they question their followers who ascertain that the answer will be sharp)

[jibu] <KIO> Aliposikia jibu [jibu] hilo, mwuaji huyo akamua (When he/she heard this answer, this killer killed him/her)

[jibu] <NUR> anashangazwa na madai ya majibu [jibu] ya Rais wa Jamhuri ya Muungano, Rais (he/she is amazed by the assertions of the answers of the President of The United Republic, the President)

{jibu} V (jibu) { answer, reply, respond, react to} AR 1009

[jibu] <ALA> kalamu na karatasi na kumjibu [jibu] Makamba kuwa yeye anampenda (the pen and the piece of paper and to answer Makamba that he likes him/her)

[jibu] <DWE> Akijibu [jibu] hoja hiyo ya Papa John (When answering this argument of Pope John)

[jibu] <KIO> Akakaa tu, asijibu [jibu] lolote wala kusalimia (He/she sat only, he/she would not answer anything nor greet)

[jibu] <MAJ> na serikali mara nyingi jibu [jibu] lao huwa chama hakina habari (and the government many times their answer is that the party does not have the news)

[jibu] <NIP> alisema Rais Mkapa wakati akijibu [jibu] swali (said President Mkapa the time when answering the question)

[jibu] fr <ALA> Akijibu [jibu] swali hilo, Waziri Mwapachu (When answering this question, Minister Mwapachu)

[jibu] fr <ALA> amefikishwa katika mahakama hiyo kujibu [jibu] shtaka la wizi (he/she has been brought to this court to answer the complaint of theft)...’

Data asilia kutoka kwa Hurskainen, 2008

Ripoti Nambari 2 ya 2008:6

4.5.1.5 Urejeleano wa vidahizo kwenye kamusi

Mfumo wa kuunda kamusi ya Kiswahili wa SALAMA-DC umejaribu kutekeleza urejeleano wa vidahizo kikompyuta kwa kutumia mbinu mbili; ya kwanza ni kuwa habari za kurejelea kidahizo kingine zitakuwa zimeshirikishwa katika uainishaji wa kimofolojia na hutumika kuundia urejeleano katik akamusi na pili ni habari zinazopatikana katika vipengele vya kileksika kwenye kamusi kwa mfano maelezo ya sehemu za tungo-funge au kwa sinonimu na maneo yaliyokaribu sinonimu. (Hurskainen, 2008).

Kwa vyovyote vile urejeleano katika kamuzi yoyote husaidia kuelewa maana zaidi na kuboresha kamusi iwe ni katika msamiati wa sinonimu au ni maelezo na habari za kietmolojia ya vidahizo vya kamusi husika.

4.5.2 Uchopoaji wa mifano ya kufafanulia maana ya vidahizo katika kamusi

Ili kamusi itimize wajibu wake ilioundiwa, lazima itoe maana ya maneno au vidahizo bila ya utata wowote. Dai hili linategemea mambo mengi hata hivyo nan i muhimu tujiulize iwapo kuna mbinu mahsusi za kutathmini suala kama hili hasa kwa kutumia teknolojia ya lugha. Kwa mujibu wa Mdee (2010:62) maana ya neno katika kamusi inachukuliwa kama sehemu muhimu zaidi katika ambayo husakwa na watuamijaji wa kamusi wanapoirejelea mara kwa mara. Kama tulivyojadili katika kipengele cha kisemantiki, neno maana tayari lina utata wa kuelezea ila Mdee 2010, alipendekeza jinsi ufanisi wa mtunga kamusi anavyofaa kupimwa mawazo yanayokaririwa na Musa 2013:121;

- a) *...kuainisha maana anuwai za vidahizo*
- b) *...kufasili vidahizo vyote vya kamusi*
- c) *...kuingiza mifano aliyoinkuu kutoka katika matini mbalimbali yenye kuthibitisha matumizi anuwai ya neno inayodhihirisha hizo maana tofauti”...*

Kwa hivyo ni dhahiri kuwa iwapo wanaleksikografia wataweza kutumia mifano tosha hasa kwa maneno ambayo si rahisi na wazi basi itakuwa hiyo ni kamusi nzuri. Je, wanaleksikografia na wataalamu wengine wanaoshughulikia kamusi kwa njia zisizoza kutumia uhandisi lugha wanajali hilo na kulitilia maanani? Hurskainen 2008:7 anadhani kuwa juhudi kidogo sana zimewekwa na wataalamu hawa na mifano ipatikanayo kwa kamusi za machapisho za kisasa ni michache mno. Pia anaongeza kuwa njia za kikompyuta zinaweza nazo kwa upande mwingine zinaweza kuibuka na kila aina ya mfano na hata mingine akawa amingi kupindukia kutokana na kongoo matini jambo ambalo pia ni changamoto tena.

Hurskainen 2008:7 ameibuka na utaratibu makini, bainifu na wa kibusara wa kuhakikisha kuwa mtaalamu anayetengeza kamusi kikompyuta ataibuka na mifano ya kufafanua maana ya maneno ambayo ni wakilishi zaidi, si mirefu kupindukia na mifupi ya kutokimu haja ya ufafanuzi unaotakiwa. Ame sema kuwa mbinu ya kuchopoa mifano kioto kutoka kwa kongoo kutumia muktadha ni changamano sana ila amechuja hoja nne za kuzingatiwa: (Hurskainen2008:8);

(a) Hatua ya kwanza ni kuzalisha orodha ya maneno yaliyopangwa kwa idadi ya utokeaji kwenye kongo kwa misingi ya mashina ya maneno na kasha kuondosha kwenye matini nambari, viwakilishi, majina maalumu (kama mahali au majina ya watu) na maneno yaliyoendelezwa vibaya. Orodha iliyobakia itasaidia kuamua ni mashina yapi yasiyohitaji mifano.

(b) Kwa kutumia ²³ *msimbo ghama* au *msimbo wa kipekee* wa kualamisha mashina katika sentensi, unda program mbili zinazoonyesha maneno (mashina ya maneno) yote yaliyotokea.

.....
²³ *msimbo ghama/wa kipekee*- Unique Code- alama mahsusi za kutambulisha kila shina katika sentensi ili kukwepa uwezekano wa kuchanganya maneno yanayohitaji kutolewa mifano

Programu moja itakuwa ni ya maneno ambayo ni nadra kupatikana na y a pili itakuwa ni yale yaliyozoeleka na yatakayochukua mifano mifupi tu na kueleweka. Hili litalinda uweza wa kutoondosha mifano ya kipekee katika neno fulani

(c) Kujaribu kufupisha sentensi ndefu zaidi kwa kuzigawanya katika mipaka ya kishazi ili angalau zibakie na mashina 40 iwapo hilo linawezekana. Kama haliwezekani unaweza kukatia popote.

(d) Mwishowe, baada ya kila shina kutambulishwa kimsimbo nakili nakili shina moja moja katika mwanzo wa mstari na kuondoa maneno/mashina mengine yote katika sentensi. Iwapo kulikuwa na mashina kumi kwa mfano katika sentensi, basi hiyo ni mifano kumi na kadhalika. Ili kukwepa tatizo la kuwa na jalada nzito zaidi lililosababishwa na mifano mingi kupindukia, fanya uteuzi nasibu wa mifano unayohitaji ili kuifupisha. Hapo utapata mifano tosha yenye miktadha yote ya matumizi kutokana na kongoo iliyotumika.

4.5.2 1 Mifano kwa kufupisha muktadha wa matumizi

Mchakato uliotumiwa na SALAMA-DC kwa mujibu wa Hurskainen 2008:8 baada ya kupata shina ktika sentensi ni ufupishaji wa mifano mirefu kupita kiasi kwa njia mbili;

a)‘...kutafuta mpaka wa kishazi na kuondosha sehemu ya sentensi nje ya mpaka.

b)...Iwapo mpaka haupo, katia au ufupishie panapofaa kwa mfano hakikisha kuwa neno lianzalo si alama ya uakifishi kwa mfano...’

4.5.2.2 Kuchagua mifano

Kwa kuwa hii ni kamusi ya kikompyuta na itahitaji mtumiaji kuchopoa kuitumia kwa kusakura mamilioni ya maneno, Hurskainen 2008 anapendekeza uondoaji wa mifano iliyoziadia kwa maneno yatokeao mara nyingin sana na kuacha ile ya maneno yatokeao mara chache mno na hufanywa hivyo kutumia uteuzi nasibu na kwa misingi ya miktadha inayotokea mara kwa mara.

4.5.2.3 Uteuzi -nasibu wa mifano uliodhibitiwa

Kwa kutumia programu ya SALAMA –DC Hurskainen 2008:9 anafafanua jinsi mfumo wa kompyuta unaounda kamusi ya kiteknolojia unavyoweza kupangwa na ukaweza kuteua kwa njia nasibu ila iliyoelekezwa mifano ya kuelezea maana ya kamusi mathalan miwili miwili kutoka kwa kila gazeti iwapo kategoria ndogo za kongoo husika ni za magazeti tofauti kama ilivyo katika mfano ufuatao alioweza kuchopoa mifano miwili miwili kwa kwa kongoo ya HCS katika magazeti ya Uhuru, Nipashe, Alasiri, Kiongozi na Nuru;

MFANO 48

‘...{andika} V (andika) {write} 627

[andika] <ALA> kuwa ni kusoma na kuandika [**andika**] na kwamba jitihada za

mwanakisomo
 [andika] <ALA> waliotapeliwa fedha zao kujitokeza kuandika [andika] taarifa zao ili waliohusika wafikishwe
 [andika] <KIO> licha ya kujifunza kusoma na kuandika [andika] vilevile alijifunza katekisimu na baadaye
 [andika] <KIO> lile la majaji walikuwa wakiandika [andika] kile alichokisema, kisha jaji
 [andika] <NIP> kuondoa hisia potofu kuwa kuandika [andika] wosia ni kuashiria au kuharakisha
 [andika] <NIP> waandishi wa habari nchini kuandika [andika] habari sahihi zinazohusu kujikinga na
 [andika] <NUR> Kondo alihoji ni vipi aliyeandika [andika] kwenye baiskeli yake ”Yesu
 [andika] <NUR> ya siku zile na kuyaandika [andika] wakati mwingine wakiandikiana barua na
 [andika] <RAI> na uwezo wa vyombo kuandika [andika] habari hizo kama vile waandishi walikuwa
 [andika] <RAI> ya wasiojua kusoma na kuandika [andika]
 [andika] <UHU> walikuwa wanajua kusoma na kuandika [andika], chini ya mpango wa Elimu
 [andika] <UHU> wengi hawana utaratibu wa kuandika [andika] wosia wakati wa uhai wao...”

Data asilia kutoka kwa Hurskainen, 2008
 Ripoti Nambari 2 ya 2008:9

4.5.2.4 Uteuzi kwa misingi ya muktadha wa idadi ya utokeaji

Kama tulivyodokezea hapo awali, mifano ya kuelezea maana za kanusi inawe kuteuliwa kwa misingi ya mara au idadi za utokeaji wa maneno kwa kuchunguza muktadah wa utokeaji huo. Hili hufanywa kwa kutafuta idadi ya kadiri ya maneno ya kushughulikia kwa kuangalia maneno yaliyo pande zote za neno kuu kama shina kutgemea kongoo. Iwapo ni maneno mawili mawili kila upande kanuni kama hiyo inaweza kuundwa kwa kongoo na majaribio mengine. Hapo tutapata ni miktadha gani unaotokea mara kwa mara na unahitaji kufafanuliwa zaidi kama ni kwa kongoo ya 1, ya 2 au ya 3 na kadhalika. Ushapata miktadha ya hiyo ya idadi ya utokeaji kutoka kwa kongoo tatu, basi utaibuka na orodha nzuri ya mifano ya kamusi yako kutegemea ukubwa uliolenga (Hurskainen 2008:10). Mfano wa 48

MFANO 49

Hatua ya 1. Neno kuu laonekana kwa ‘dirisha ‘hilo’:
 (a) ...tangu yalipoanza [anza] machafuko ya
 Hatua ya 2. Mfano wa urefu wa tungo na neno kuu:
 (8) ...[anza] <DWE> wa Israel wameuawa tangu yalipoanza [anza] machafuko ya Intifada

Data asilia kutoka kwa Hurskainen, 2008
 Ripoti Nambari 2 ya 2008:10

4.5.3 Kuunganisha vijisehemu ili kuunda kamusi

Hatimaye vipande vitatu vilivyofafanuliwa hapo juu vinaweza kuunganishwa na kuwa zao ambalo sasa ni kamusi, yaani jalada lenye vidahizo au vipashio vya kileksika (msamiati), pili ni faili yenye mifano iliyoteuliwa kwa njia nasibu kulingana na muktadha na tatu ni failiau jalada lenye mifano ya

miktadha ya utokeaji wa idadi za maneno. Majalada hayo matatu yakiunganishwa pamoja huku neno kuu linatangulia, likafuatwa na mifano iliyoteuliwa kwa njia nasibu na hatimaye mifano ya miktadha ya mara ya utokeaji na kadhalika (Hurskainen 2008). Mfano ufuatao ni dondoo kutoka kwa kamusi ambayo aliunda kwa njia hii kutokana ka sehemu za kongoo ya HCS kategoria ndogo za magazeti ya Alasiri, Kiongozi. Majira, Raia, nipashe, Nuru na tovuti ya habari Deutsch Welle kutoka Ujerumani.

MFANO 50

‘... {anza} V (anza) { begin, start } 4289
 [anza] <ALA> Fresh ya nao walionekana kuanza [anza] kutafuta usafiri mwingine, baada ya
 [anza] <ALA> kamati ya chama hicho akaanza [anza] kumrushia madongo mwenyekiti wake mbele
 ya
 [anza] <DWE> na hofu yoyote na kuanza [anza] kwenda makazini kama kawaida hii
 [anza] <DWE> serikali mpya, Ujerumani imeshaanza [anza] bila kuchelewa kuitolea nchi yake
 [anza] <KIO> hapo alinyanyuka na kuanza [anza] kunipiga na kwendea kisu alichokuwa
 [anza] <KIO> anaonyesha mkono huo na kuanza [anza] kulia) hapo alinikamata hadi
 [anza] <MAJ> Acb ilianza [anza] kufanya shughuli zake hapa nchini
 [anza] <MAJ> kulala, na mchezo ukaanzia [anza] hapo"...
 [anza] <NIP> "africa" One kuanza [anza] safari nchini karibuni,
 [anza] <NIP> cha mjini Arusha, kitaanza [anza] uzalishaji wa dawa za maji
 [anza] <NUR> "mimi" naanza [anza] kwa msemo wa hapo juu ya kwamba
 [anza] <NUR> Aidha, lilipitishwa azimio kuwa kuanzia
 [anza] <RAI> "Sisi tunataka kuanzia [anza] hapo tujue kama mali inayotangazwa
 [anza] <RAI> maalum, Naila Jiddawi kuanza [anza] kujadiliwa jana
 [anza] <UHU> Adhabu ya kufungiwa inaanza [anza] Januari 20, 2003",
 [anza] fr <DWE> Waisraeli wameuawa, tangu yalipoanza [anza] machafuko ya Wapalestina,
 mwishoni
 [anza] fr <DWE> Waisraeli 71 wameuawa tangu yalipoanza [anza] machafuko ya Wapalestina,
 Intifada...’
 [anza] fr <DWE> itikadi kali, tangu yalipoanza [anza] machafuko ya Wapalestina mwezi wa
 [anza] fr <DWE> kabisa mjini Jerusalem tangu yalipoanza [anza] machafuko ya Wapalestina zaidi ya
 miezi
 [anza] fr <DWE> miongoni mwa Wapalestina, tangu yalipoanza [anza] machafuko ya Wapalestina
 mwezi wa
 [anza] fr <DWE> wa Israel wameuawa tangu yalipoanza [anza] machafuko ya Intifada.
 [anza] fr <DWE> wa itikadi kali tangu yalipoanza [anza] machafuko ya Intifada.

Data asilia kutoka kwa Hurskainen, 2008

Ripoti Nambari 2 ya 2008:11

Kutoka na kitenzi –anza- ambalo ni shina la kitenzi hicho bila mnyambuliko, tumeibuka na kidahizo cha kamusi iliyojengwa kikompyuta na programu ya SALAMA-DC. Kama tulivyodokezea ni kuwa kila kitu kimefanywa kioto ikiwepo tafsi ya neno anza kwa Kiingereza *'begin, start'*. Ikumbukwe tafsiri hii inaweza kuhitaji kuhaririwa kidogo iwe iwakilisha vizuri Kiingereza sahihi. Baada ya mifano miwili miwili ya kuchopolewa kwa njia nasibu, inafuata ile ya idadi ya utokeaji kimuktadha katika DWE ya ‘dirisha la maneno mawili mawili kila upande yaani (2+2). Hapa kulikuwa na nafasi ya kubadilisha idadi ya maneno kwa muktadha kikompyuta na

kupata mifano anuwai bila kuongeza idadi ya mifano (Hurskainen 2008). Tayari tunayo kamusi ya Kiswahili iliyojengwa kioto kwa kutumia msamiati wa kongoo na programu za SALAMA za teknolojia ya lugha.

Iwapo mifano ni haba katika Kongoo kama katika 50 hapa chini, basi tunapata mifano michache;

MFANO 51

‘... (10) {abudia} V (abudu) { worship } APPL 4
[abudia] <ALA> yanafanywa kwenye nyumba za kuabudia [abudia].
[abudia] <BIB> Yerusalemu ni mahali patupasapo kuabudia [abudia].
[abudia] <NIP> miradi ya nyumba za kuabudia [abudia].
[abudia] <NUR> Kwani sisi tunaabudia [abudia] kiwanda au mameneja wake mpaka...’
Data asilia kutoka kwa Hurskainen, 2008
Ripoti Nambari 2 ya 2008:12

4.5.4 Kutafsiriwa kwa mifano katika kamusi iliyoundwa

Kwa kuwa sasa tunayo kamusi tayari, tunaweza kuichezea tutakavyo na kuzalisha matokeo kadhaa kutokana na kamusi hii. Tafsiri-mashine ni mojawapo ya mbinu na uwezo wa SALAMA wa kushughulikia lugha kikompyuta. Kama tulivyosema awali ni kuwa baadhi ya program zinaingiliana na hivyo pia katika ni Hurskainen 2008:12 anatumkumbusha kuwa ni vidahizo abudu na *anza* na *anzia* vina chukuliwa na mfumo huu kumaanisha kitu kimoja na hivyo ni bora kufanya uondoshaji wa utata wa kisemantiki. Mfano wa 51 unaonyesha baadhi ya mifano ya tafsiri ya dondoo la kamusi kwa Kiingereza;

MFANO 52

‘... {anza} V (anza) { begin, start }
[anza] <ALA> kamati ya chama hicho akaanza [anza] kumrushia madongo mwenyekiti wake mbele ya (the committee of this party began to throw dirt on its chairman in front of)
.....
[anza] <DWE> serikali mpya, Ujerumani imeshaanza [anza] bila kuchelewa kuitolea nchi yake (the new government, Germany has begun without being late to give to its country)
.....
[anza] <KIO> hapo alinyanyuka na kuanza [anza] kunipiga na kwendea kisu alichokuwa (on that moment he/she rose up and began to hit me and to go the knife which he/she had)
...
[anza] <MAJ> ACB ilianza [anza] kufanya shughuli zake hapa nchini (ACB began to do its activities here in the country)....
[anza] <NIP> "Africa" One kuanza [anza] safari nchini karibuni, ("Africa" One to begin the journey in the country soon,).....[
[anza] <NUR> "mimi" naanza [anza] kwa msemo wa hapo juu ya kwamba ("I" begin with the locution above that).....
[anza] <RAI> "Sisi tunataka kuanzia [anza] hapo tujue kama mali inayotangazwa ("We want to begin here so that we would know if the ostensible property)....
[[anza] <UHU> Adhabu ya kufungiwa inaanza [anza] Januari 20, 2003", (The punishment of being jailed begins January 20, 2003",)

Data asilia kutoka kwa Hurskainen, 2008
Ripoti Nambari 2 ya 2008:13

Kwa mifano michache ya tafsiri tuyoitoa hapo juu katika 51, ifahamike kuwa hii ni tafsiri-mashine ya sampuli ya matini tu na si ya sentensi nzima, ambyo ingekaa bora zaidi kisarufi, ila bado, mfumo huu wa Tafsiri-oto umejaribu tafsiri ya kimsingi mno ambayo inarekebishika kwa kuhaririwa. Tutaangalia mifani michache ya jinsi SALAMA-DC pia inaweza kuunda kamusi ya fungo-tunge kwa kuzitenga, kuzipa mifano, kuzitafsiri kama kamusi tu ya neno moja.

MFANO 53

‘...({piga} V (piga) { hit, beat }
 {piga_pasi} V IDIOM-V { iron clothes } 3
 [piga_pasi] <ALA> hapo mtu asipike wala asipige [piga_asi] pasi" (here man should neither cook nor iron clothes")
 [piga_asi] <NIP> mara baada ya marehemu kumaliza kupiga [piga_asi] pasi nguo zake kisha wakamweka chini ya (immediately after the late to finish to iron his clothes then they placed him/her under)
 {piga_picha} V IDIOM-V { photograph } 40
 [piga_picha] <ALA> Ikulu kunywa chai na kupiga [piga_picha] picha na Rais Mkapa (the State House to drink tea and to photograph with President Mkapa)
 [piga_picha] <DWE> kutoka Ujerumani, walijitahidi kupiga [piga_picha] picha za ukumbusho na kiongozi wao (from Germany, they made an effort to photograph the Commemoration and their leader) {piga_ramli} V IDIOM-V { divine } 4
 [piga_ramli] <KIO> anakwenda kwa mganga ili kupiga [piga_ramli] ramli na kuongeza imani za ushirikina (he/she goes to the medical person in order to divine and to increase the faith of superstition)
 [piga_ramli] <KIO> kwenda kwa mganga wa kupiga [piga_ramli] ramli, hujui kuwa imani ya (going to the medical person of divining, you do not know that the faith of) ...’

Data asilia kutoka kwa Hurskainen, 2008
 Ripoti Nambari 2 ya 2008:13

Kama tulivyofafanua jinsi SALAMA-Dc inavyoweza kuunda mifano kutokana na vidahizo vya kihomonimu kama ifuatavyo;

MFANO 54

“... {jibu} N 5/6 { answer, reply, response } AR
 [jibu] <DWE> wanahoji wafuasi wake wanaohakikisha jibu [jibu] litakuwa kali (they question their followers who ascertain that the answer will be sharp)
 [jibu] <KIO> Aliposikia jibu [jibu] hilo, mwaaji huyo akamua (When he/she heard this answer, this killer killed him/her)
 [jibu] <NUR> anashangazwa na madai ya majibu [jibu] ya Rais wa Jamhuri ya Muungano, Rais (he/she is amazed by the assertions of the answers of the President of The United Republic, the President)
 {jibu} V (jibu) { answer, reply, respond, react to } AR 1009
 [jibu] <ALA> kalamu na karatasi na kumjibu [jibu] Makamba kuwa yeye anampenda (the pen and the piece of paper and to answer Makamba that he likes him/her)
 [jibu] <DWE> Akijibu [jibu] hoja hiyo ya Papa John (When answering this argument of Pope John)

 [jibu] <NIP> alisema Rais Mkapa wakati akijibu [jibu] swali (said President Mkapa the time when answering the question)
 [jibu] fr <ALA> Akijibu [jibu] swali hilo, Waziri Mwapachu (When answering this question, Minister Mwapachu)...”

Data asilia kutoka kwa Hurskainen, 2008
 Ripoti Nambari 2 ya 2008:13

SALAMA-DC inaweza kutafuta na kutafsiri methali za Kiswahili kwa mfano katika Kongoo 2: kwenye mfano ufuatao;

- {... {Asiyefunzwa na mamaye, hufunzwa na ulimwengu} (Who is not taught by one's? mother is taught by the world) 1
 {Bendera inafuata upepo} (The flag follows the wind) 1
 {Hujafa, hujaumbika} (You are not yet dead, so you are not yet fully created) 2
 {Haba na haba hujaza kibaba} (Little and little fills up the tin) 1
 {Hasira hasara} (Anger brings damage) 1
 {Kutangulia si kufika} (To go first is not to arrive) 2
 {Kutoa ni moyo, usambe ni utajiri} (Giving is from heart, do not say it is from richness) 1
 {Mgeni njoo mwenyeji apone} (Let the guest come and the host be rescued) 1
 {Mkono mtupu haulambwi} (The empty hand is not licked) 1
 {Mla nawe hafi nawe ila mzaliwa nawe} (The one who eats with you does not die with you like the one who was born with you) 1
 {Msafiri kafiri} (The traveller is an unbeliever) 1
 {Bendera hufuata upepo} (The flag follows the wind) 3
 {Penye wengi hapaharibiki neno} (Where there are many people, the word is not misused) 1
 {Tamaa mbele mauti nyuma} (Greediness first, death later) ...'

Data asilia kutoka kwa Hurskainen, 2008
 Ripoti Nambari 2 ya 2008:14

4.5.5 Utathmini wa SALAMA-DC kutumia kongoo-matini tatu za Kiswahili

Baada ya kuijaribu programu ya SALAMA-DC na kongoo 3 mambo kadhaa yalidhihirika.

Kongoo 1 ilikuwa na maneno (148,700 words) ambayo unaundwa na vitatu vitano vya fasihi bunilizi vya E. Kezilahabi. Kongoo ya 2 (maneno milioni 3) ya matini kuhusu habari na Kongoo ya 3 ikiwa na maneno milioni 20 kama tulivyotaja katika sura ya pili. Majaribio haya yalifanyiwa kwenye kipakatalishi cha uwezo wa RAM 512KB. Data ifuatayo ilishuhudiwa na Hurskainen2008.

MFANO 56

Jedwali: 3- Sifa za Kamusi za Majaribio zilioundwa kutoka kwa Kongoo 3

	Corpus 1	Corpus 2	Corpus 3
	Maneno 148,700	Maneno milioni 3	Maneno milioni 20
Vidahizo (neno moja)	5,961	15,189	25,576
Vidahizo (tungo-funge)	1,001	1,489	2,381
Kurejeleana (tungo-funge)	1,117	1,495	2634
Kurejeleana (sinonimu)	2,085	2,945	5,212
Mifano katika muktadha(yaani iliyoteuliwa kiholela)	10,112	105,818	173,082
Mifano katika muktadha (kwa msingi wa uchanganuzi maneno kiidadi)	1,883	31,258	11,048

(Hurskainen, 2008- Ripoti Namba.2-2008: Ukurasa wa 15-16) (Tafsiri yetu)

Wakati kamusi iliyoungwa na SALAMA-DC ilijaribiwa na kongoo matini tatu hizi, Kongoo 1 na 2 ziliweza zilipita awamu zote bila tatizo lolote ila uchakataji wa Kongoo 3 ulihitaji kugawanya programu ili iweza kuchopoa mifano yote ya sentensi kwa sababu ya nafasi ya hifadhi. Hurskainen, 2008 anaongezea kuwa kipakatalishi cha siku hizo kitatekeleza wajibu huo bila wasi wasi wowote.

Kwa mujibu wa Hurskainen 2008, majaribio haya yalipata Kongoo ya kwanza na ya pili zikiwa hazina changamoto za kuchopoa mifano kwa muktadha wa idadi ya utokeaji. Kongoo ya 3 ilihitaji zizuizi na uchujaji kwa kuwa mifano mingine ilifikia hata mamia (Hurskainen 2008:16) kwa hivyo baada ya uteuzi nasibu ilibakia na mifano kama 3 mitatu hivi.

Kwa kuangalia mchakato huu ambao unaishia kuwa na kamusi ya aina hii ikiwa hata na tafsiri ya maana na mifano kwa Kiingereza. Pia mchakato wote huu wa kutoka matini asilia hadi kuwa na kamusi ya aina huwa ni wa kioto na hakuna mahali pana mkono wa mtu katika uhariri. Shughuli kama hii huchukua saa chache na ingalikuwa ni uundaji wa kamusi wa kawaida ingechukua miezi na hata miaka.

4.5.6 Jinsi ya kutumia kamusi hii

SALAMA-DC hutoa kamusi ikiwa katika muundo wa matini kama jalada moja na baadhi ya watumaji hupenda ikiwa hivi kwa kuwa wanaweza kutumia programu za kuchopoa data watakazo. Uhifadhi wake unaweza pia kwa kanzi-data za aina mbalimbali kwa wale wasio na uzoefu mkubwa wa kutumia kompyuta (Hurskainen 2008:17).

4.5.7 Matumizi ya SALAMA katika tathmini ya Kamusi tano za Kiswahili

Kwa mujibu wa maelezo tuliyotoa katika sehemu hii, ni wazi kuwa wanaleksikografia wana mengi ya kujifunza kutokana na mifumo ya kuunda kamusi na kutafiti masuala ya uundaji kamusi kutumia kwa kutumia mbinu za teknolojia ya lugha hasa za kikongoo. Baada ya kufahamu jinsi ya kuunda kamusi, ni vyema pia tutalii jinsi ya kutathmini ubora wake. Tajriba ya mradi wa SALAMA itatuongoza katika shughuli hii. Hurskainen 2002:283 anafikiri kuwa wataalamu wa kamusi hata wanaoitwa wanaleksikografia wa Kiswahili siku hizi hawana mbinu za kuridhisha za kutathmini kamusi inavyotakiwa na anahisi kuwa badala ya kudadisi na kutathmini kamusi za njia za kutegemeka na zilizo sahihi, huwa wataalamu hawa na watafiti wengine wanaohusika wanaishia tu toa maoni. Maoni haya hutofautiana kwa kuwa mara nyingi hayana njia ya hakika ya kuibuka nayo. Sehemu hii itaangalia jinsi teknolojia ya lugha ya SALAMA ilivyotumika katika utathmini wa kamusi tano kwa njia za kikompyuta. Kamusi hizi ni Kamusi ya Kiswahili Sanifu-KKK (TUKI)

iliyosemekana kuwa ina vidahizo 14,288 ila kamusi yenyewe ikapatikana na vidahizo 20,000 na maneno 50,000, Kamusi ya Maana na Matumizi (OUP) ilisemekana ina vidahizo 8,057 ila baada ya tathmini ya SALAMA ikakutikana na vidahizo 9,000 na mifano ya matumizi 12,000, Modern Swahili - Modern English Dictionary (MS-tryck) ilitarajiwa kuwa na vidahizo 10,461 lakini kamusi kamusi yenyewe ikaishia kupatikana na maneno 13,300 na mifano ya matumizi 3,000, Kamusi ya Kiswahili - Kiingereza (TUKI) ilidhaniwa itakuwa na vidahizo 14,533 na hatimaye ikapatikana na zaidi ya vidahizo 30,000, na kamusi ya tano ni Swahili - Suomi - Swahili –sanakirja (SKS) (Yaani kamusi ya Kiswahili kwenda Kifini na Kifini kurudi Kiswahili ilifikiriwa itakuwa na vidahizo 11,500 ila kamusi yenyewe ilikutikana na kama vidahizo 10,000 (Hurskainen, 2002:284).

Programu za SALAMA ambazo zinaainisha maneno ya Kiswahili na kuteua lemma (mashina ya maneno yanayoweza kuwa vidahizo) katika matini. Kongoo tatu, kama tulivyodokezea kuhusu utafiti huu, zilitumika na ufanisi wa kila kamusi ulitathminiwa kwa kutumia aina tatu za kongoo matini. Utafiti huu wa Hurskainen 2002 unaonyesha uzuri na upungufu wa kila kamusi na kupendekeza njia mwafaka za kufidia upungufu huo. Ili kufanikisha hilo, kwanza ilikuwa ni sharti kamusi itimize masharti manne muhimu (Hurskainen 2002:286); nayo ni

- i) Kamusi kuwa na yote yaliyo kwenye kongoo kama vidahizo vyake,
- ii) Kila kidahizo kifafanuliwe vya kutosha kuonyesha ngeli iwapo ni nomino na minyambuliko iawapo ni vitenzi,
- iii) Iwe wazi kama vivumishi vinanyambulika au la
- iv) Kamusi isiwe na vidahizo ambavyo havijafafanuliwa na kuonyesha mifano

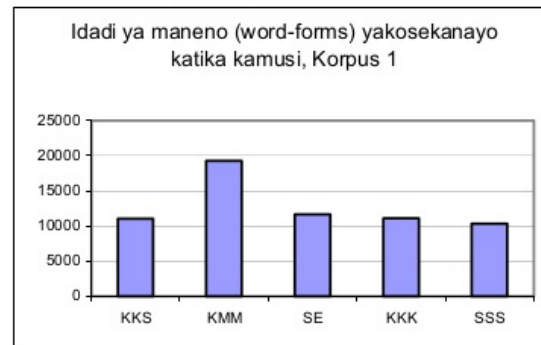
Kwa kutumia matini kutokana na kongoo tatu zilizotajwa katika 3.6.1 hapo juu, Hurskainen 2002) alilenga kutafuta maneno yapatikanayo katika matini kongoo ila hayapatikani kama vidahizo katika kamusi na pili, kubainisha maneno yaliyo kama kidahizo kwenye kamusi, lakini hayatokei kwenye kongoo-matini. Tatu, alilenga kulinganisha kamusi zote tano katika masuala mawili yaliyotajwa katika 1 na 2 hapo juu na kudhihirisha ubora na upungufu wa kila kamusi na hatimaye kulinganisha Kamusi ya Kiswahili Sanifu na Kamusi ya Kiswahili-Kiingereza kwa kuona maendeleo yaliyofanywa kati ya miaka 1981 na 2001 katika kuchagua vidahizo.

Mifano ifuatayo ya data za utathmini huu imedondolewa kwa kuzingatia vigezo vya malengo haya ili kuonyesha jinsi kutafiti kwa ubora, upungufu, mkao na mabadiliko au mapito ya kamusi yanavyoweza kutafitiwa kutumia teknolojia ya lugha ya SALAMA:

MFANO 57

Jedwali 1. Idadi ya maneno yakosekanayo katika makamusi, Korpus 1.

KKS	11054
KMM	19245
SE	11680
KKK	11101
SSS	10364

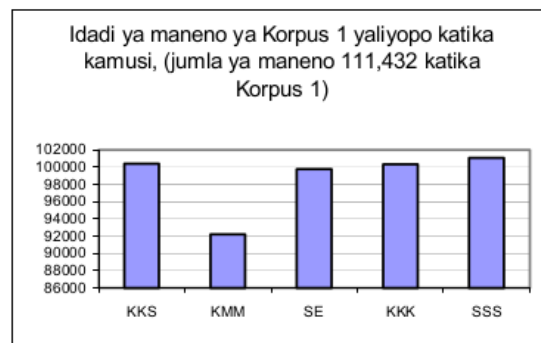


Data asilia kutoka kwa Hurskainen 2002:289

MFANO 58

Jedwali 2. Jumla ya miundo ya maneno yaliyogunduliwa na makamusi katika Korpus 1.

KKS	100378
KMM	92187
SE	99752
KKK	100331
SSS	101068

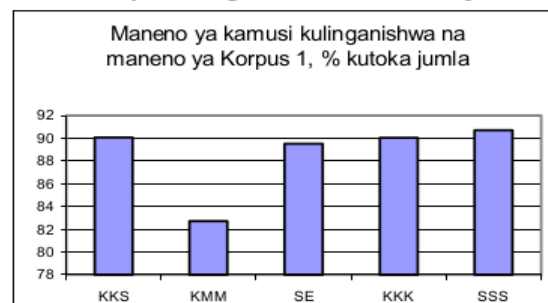


Data asilia kutoka kwa Hurskainen 2002:289

MFANO 59

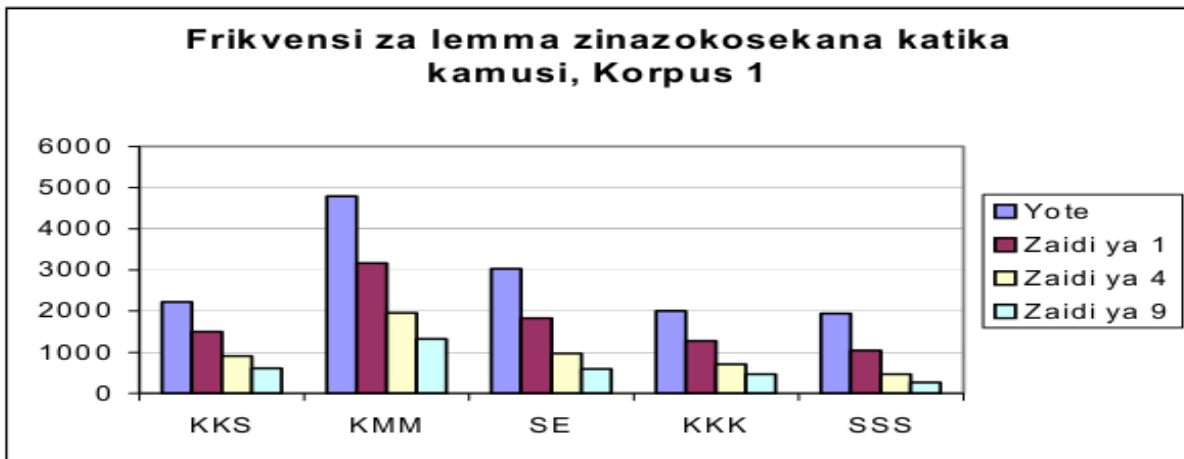
Jedwali 3. Maneno ya kamusi kulinganishwa na maneno ya Korpus 1, % kutoka jumla.

KKS	90,08
KMM	82,73
SE	89,52
KKK	90,04
SSS	90,7
TOTAL	100



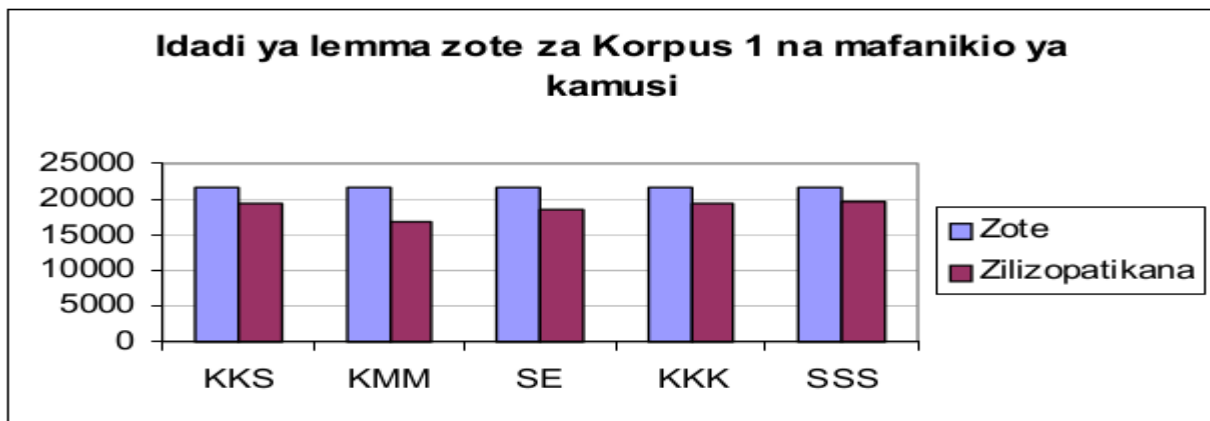
Data asilia kutoka kwa (Hurskainen 2002:289)

MFANO 60



Data asilia kutoka kwa (Hurskainen 2002:290)

MFANO 61

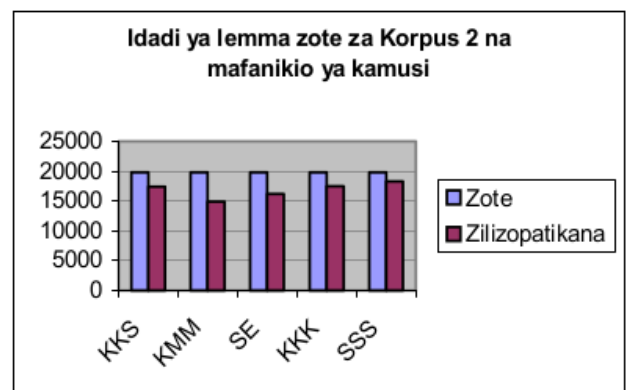


Data asilia kutoka kwa (Hurskainen 2002:291)

MFANO 62

Jedwali 7. Idadi ya lemma zote za Korpus 2 na mafanikio ya kamusi.

	Zote	Zilizopatikana	%
KKS	19799	17379	87.8
KMM	19799	14875	75.1
SE	19799	16152	81.6
KKK	19799	17488	88.3
SSS	19799	18287	92.4

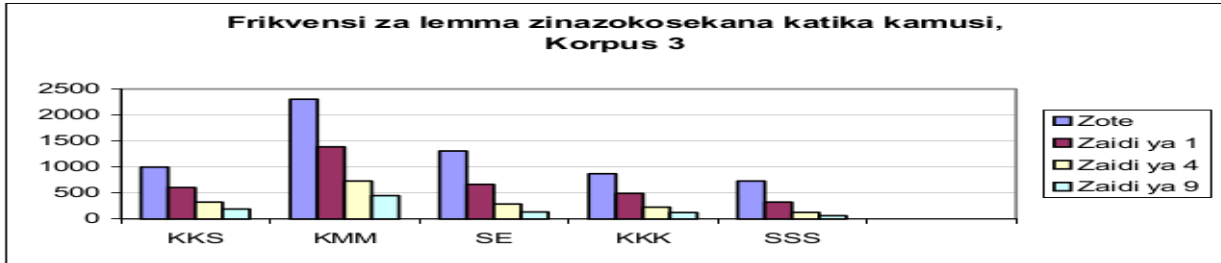


Data asilia kutoka kwa (Hurskainen 2002:292)

MFANO63

Jedwali 8. Frikvensi za lemma zinazokosekana katika kamusi, Korpus 3.

	Zote	Zaidi ya 1	Zaidi ya 4	Zaidi ya 9
KKS	1001	605	325	193
KMM	2304	1387	731	449
SE	1309	665	287	136
KKK	871	494	229	125
SSS	731	326	128	64

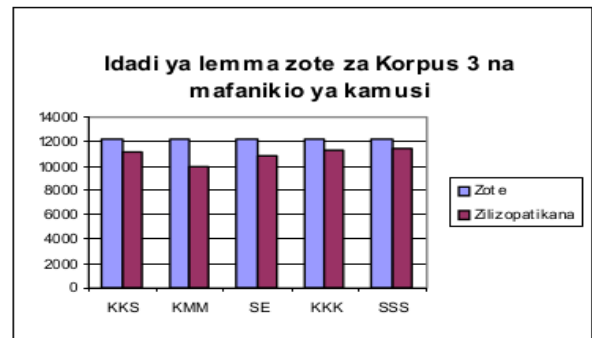


Data asilia kutoka kwa (Hurskainen, 2002:293)

MFANO 64

Jedwali 9. Idadi ya lemma zote za Korpus 3 na mafanikio ya kamusi.

	Zote	Zilizopatikana	%
KKS	12209	11208	91.8
KMM	12209	9905	81.1
SE	12209	10900	89.3
KKK	12209	11338	92.9
SSS	12209	11478	94.0

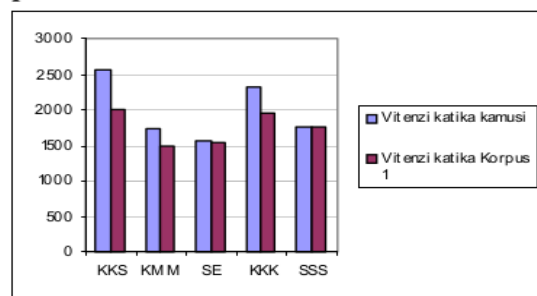


Data asilia kutoka kwa (Hurskainen 2002:293)

MFANO 65

Jedwali 10. Vitenzi katika kamusi na katika Korpus 1.

	Vitenzi katika kamusi	Vitenzi katika Korpus 1
KKS	2560	1991
KMM	1742	1500
SE	1557	1540
KKK	2305	1948
SSS	1751	1748

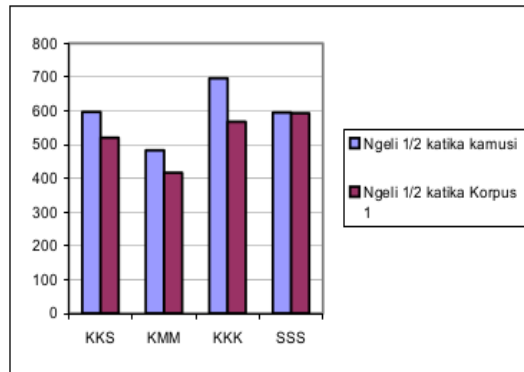


Data asilia kutoka kwa (Hurskainen 2002:295)

MFANO 66

Jedwali 11. Ngeli 1/2 katika kamusi na katika Korpus 1.

	Ngeli 1/2 katika kamusi	Ngeli 1/2 katika Korpus 1
KKS	597	520
KMM	483	417
KKK	696	568
SSS	595	593

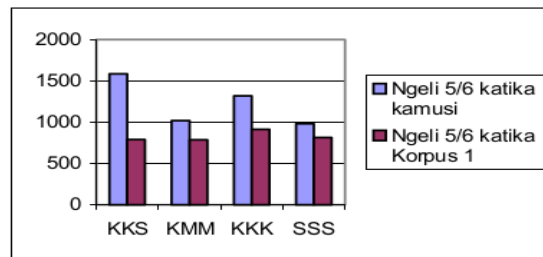


Data asilia kutoka kwa (Hurskainen 2002:296)

MFANO 67

Jedwali 13. Ngeli 5/6 katika kamusi na katika Korpus 1

	Ngeli 5/6 katika kamusi	Ngeli 5/6 katika Korpus 1
KKS	1584	788
KMM	1020	786
KKK	1319	914
SSS	981	815

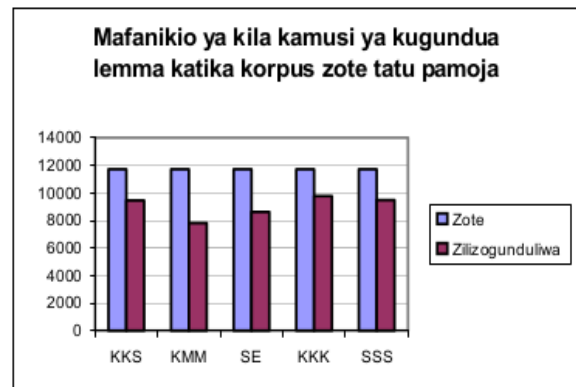


Data asilia kutoka kwa (Hurskainen 2002:296)

MFANO 68

Jedwali 16. Mafanikio ya kila kamusi katika kugundua lemma katika korpus zote tatu pamoja.

	Zote	Zilizogunduliwa	%
KKS	11713	9441	80.6
KMM	11713	7808	66.7
SE	11713	8613	73.5
KKK	11713	9746	83.2
SSS	11713	9475	80.9

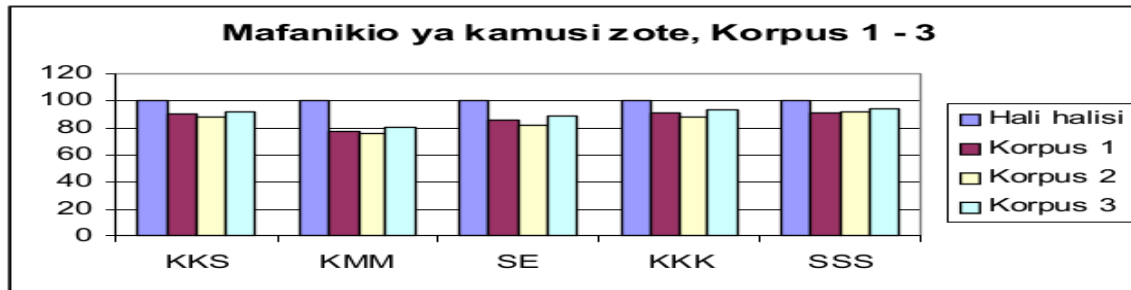


Data asilia kutoka kwa (Hurskainen 2002:298)

MFANO69

Jedwali 17. Mafanikio ya kamusi zote, Korpus 1-3.

	Hali halisi	Korpus 1	Korpus 2	Korpus 3
KKS	100	89,7	87,7	91,8
KMM	100	77,7	75,1	81,1
SE	100	85,9	81,6	89,3
KKK	100	90,7	88,3	92,9
SSS	100	91	92,4	94



Data asilia kutoka kwa (Hurskainen 2002:298)

4.5.8 Uchanganuzi wa data za uundaji na utathmini wa kamusi kikompyuta

Kwa mujibu wa data zilizowasilishwa hapa kuhusu tathmini ya kamusi tano za Kiswahili kutumia kongoo na program za SALAMA, imeibuka kuwa a) kamusi zote tano zina upungufu katika kuingiza vidahizo kwa kutimiza mahitaji ya matini zilizotumiwa ambazo ni kongoo tatu zilizobainishwa hapo awali.

(2) Utathmini huu ulipata kuwa kulikuwa na tofauti za maana kati ya kamusi hizo. Mathalan kamusi ya KMM hasa ilipatikana kutofautiana na kamusi nyinginezo na matokeo yake yalisemekana ni hafifu na kupendekezwa iangaliwe upya kwa misingi hiyo. (Hurskainen 2002). Hii ni dhahiri kuwa matumizi ya mbinu za SALAMA yaliweza kuibuka na masuala mahsusi kuhusu hali ya kamusi na kupendekeza suluhu. Isitoshe, tathmini hii na mifano iliyotolewa inasadia kupuuzilia mbali imani kuwa ukubwa wa kamusi ndiyo ubora wake. Ni ajabu kuwa Hurskainen 2002, anasema kuwa si kweli kuwa kamusi yenye maneno zaidi ndiyo inayofanikiwa vizuri ka sababu kwa kutumia programu za SALAM-DC anakuta kuwa kamusi ya KKS na KKK zilizo kuwa na vidahizo vingi kuliko kamusi nyingine, ila maajabu nikuwa kamusi ya SSS iliyo na vidahizo 11.000 pekee ilizishinda kamusi zote katika kila aina ya tathmini. Hii ni njia rahisi, ya hakika nay a kisayansi katika kutafitia michakato ya kileksikografia. Isitoshe kamusi KKK na KKS ni kamusi ambazo zina maneno ya aina fulani kuliko kamusi nyingine na maneno hayo yalionekana kuwa hayakutumiwa katika matini asilani.

Fauka ya hayo tunaona kuwa kamusi ya KKK na KKS zote mbili zilikuwa karibu sawa katika kutambua miundo ya maneno ya magazeti, lakini ifikapo utokeaji wa ya mashina ya maneno (lemma) utafahamu kuwa kamusi ya KKK ilikuwa nzuri kwa kiasi fulani (Hurskainen 2002).

Isitoshe, hali bora ya kamusi ya SSS inaonyesha kwamba mbinu za kikompyuta zinaweza kuboresha kamusi kwa kiasi kikubwa bila ya kutegemea ukubwa wake. Pia mbinu za kikompyuta zinaweza kutumiwa kwa kuonyesha kila neno linalopungua katika kamusi, na pia kuonyesha kila kidahizo ambacho hakitumiwi katika matini au kongoo iliyotumia kutathminia kamusi husika. Pia frikvensi au mara ya utokeaji wa maneno hayo zinapatikana na hivyo kuwa njia bora za kutegemewa na zzinazoharakisha na kusaidia mchakato wa kuunda kamusi. (Hurskainen 2002).

Mfumo wa SALAMA-DC unaweza kupokea makala ya aina yoyote na kuunda kamusi kuzingatia kongoo.

Kwa kuhitimisha, ni bora ijulikane kuwa licha ya kuwa kazi hii ya uundaji kamusi kuwa inafanywa kwa njia ya kikompyuta ipo haja ya kupitiwa na kurekebisha kidogo na binadamu ambao ni mchango anaouita Hurskainen (2008) sehemu kidogo sana ya kazi inayofanywa na wanaotumia mikono kuunda kompyuta.

Kwa kugemea mkabala wa nadharia ya kongoo, utaona kuwa mchakato wa kuunda kamusi kukompyuta una manufaa mengi sana kuliko kutumia njia za kimapokeo za kiungamuzi ukiwemo urahisishaji kazi na uwezo wa kusaka maneno kwenye kamusi kikompyuta. Kama asemavyo (Hurskainen 2008a:17), utathmini wa kutumia kongoo ya tatu ulipunguza mifano ya kamusi husika kwa 2% na bado kamusi iliyozaliwa na mchakato huu ilikuwa kubwa kuliko ile asilia kwa karibu mara ishirini, na hivyo haina haja ya kuchapishwa kwa karatasi, bora kuwe na mfumo wa kusaka maana kwa kompyuta.

4.6 Teknolojia ya Lugha, Tafsiri-Mashine na Utafiti katika Kiswahili

4.6.0. Utanngulizi

Tafsiri-mashine ambayo ni mbinu ya kutafsiri lugha moja hadi Lugha Chanzi (LC) hadi lugha nyingine –lugha lengwa (LL) kutumia kompyuta. Tafsiri ni eneo muhimu katika utafiti wa lugha na limeshughulikiwa na waalamu kwa maelfu kwa miaka mingi (Hurskainen 2005:120) ili kuwezesha mawasiliano kati ya wazungumzaji mbali mbali. Kwa mujibu wa Hurskainen aina mbili za tafsiri - mashine yenye mtazamo wa kikongoo ambazo ni ile ya kutumia kongoo sambamba ambapo matini moja hutoa katika kongoo moja kwa lugha mbili husika na kupata tafsiri kwa kutafuta na kulinganisha mifuatano ya maneno ambayo anasema ina changamoto na mapungufu mengi. Njia ya

pili ni ile ya kutumia sarufi ya lugha katika uundaji wa kanuni za tafsiri baada ya kuainisha maneno kimofolojia, kuondosha utata wa maana na hatimaye kupata tafsiri kutoka kwa muktadha husika.

Hurskainen 2005, anaona kuwa mfumo huu wa pili ambao ulitumiwa katika SALAMA, ndio bora zaidi ila pia una changamoto zake na katika hatua za kuzitatua ndipo tutapata mbinu za kutafitia kipengelee hiki cha lugha ya Kiswahili kutumia teknolojia ya lugha. Katika sehemu hii tutapata mifano kadhaa na kutathmini nafasi yake katika utafiti wa Kiswahili kama ilivyotumika katika SALAMA

4.6.1. Tafsiri-mashine na vivumishi katika Kiswahili kwenda Kiingereza

Sehemu hii itahusika na hali za vitenzi vya Kiswahili ambavyo Hurskainen anasema kuwa ni kundi la maneno katika Kiswahili ambayo ni machache sana. Katika tafsiri-mashine ya Kiswahili kwenda Kiingereza huwa kunashuhudiwa matatizo ya kulinganisha vivumishi kwa kuwa vya Kiswahili ni vichache mno. Katika SALAMA mbinu kadhaa zilitumika ili kufidia upungufu huu kama ilivyo katika mifano ifuatayo;

MFANO 70

Tafsiri-mashine ya Vivumishi vya Kiswahili vyenyewe kwa Kiingereza huwa haina utata mwingi kwa kuwa vivumishi vingi vya Kiswahili vina visawe vya Kiingereza na ni nadra sana kuwe na utata wa kisemantiki. K.m. –zuri (good) na kadhalika.

MFANO 71

‘..."<chakula>" "chakula" N 7/8-SG { the } { food } MASS
"<kizuri>" "zuri" ADJ A-INFL 7-SG { good }...’

Data asilia kutoka kwa (Hurskainen 2014:2) Ripoti namba:15

Hata hivyo kuna tafsiri-mashine za vivumishi vya aina tofauti ambayo huhitaji uondoshaji wa utata wa maana ili kupata tafsiri sahihi katika sentensi au katika tungo lolote. Mfano wa 71 unawasilisha jinsi Hurskainen anavyoshughulikia tafsiri zenye vivumishi ambavyo si vya moja kwa moja kama mfano 71 ila ni tungo-funge. Mfano wa 72 unaonyesha kivumishi cha aina ya uhusika milikishi kama katika-*ruhusa ya kudumu*.

MFANO 72

‘..."<ruhusa>"
"ruhusa" N 9/10-SG { the } { permission }
"<ya>"
"ya" GEN-CON 9-SG { of }
"<kudumu>"

"dumu" N 15-SG z [dumu] { last , be permanent } SV

Baada ya kuondolewa utata tafsiri hii itaishilia kutenga tungo-funge na kuitafsiri kwa usahihi kama

"<ruhusa>"
"ruhusa" N 9/10-SG { the } { permission }
"<ya_kudumu>"
"ya_dumu" ADJ {permanent}

Matokeo ya tafsiri sahihi baada ya mhakato huo huwa ; *The permanent permission*

Data asilia kutoka kwa (Hurskainen 2014:5) Ripoti namba:15

MFANO 72

Katika kushughulikia tafsiri-mashine iliyo na kivumishi cha Kiswahili kilichoundwa na vitenzi vyenye viambishi rejeshi, Hurkainen 2014:6 anatoa mifano ya kuonyesha kuwa tafsiri yake katika Kiingereza huishia kuwa ni kitenzi nomino mathalan; hali ya hewa inayobadilika ikitafsiriwa moja kwa moja itakuwa : *the condition of air that changes* > ila mfano huu utaonyesha ni vipi tunafikia tafsiri sahihi ya kupata *changing weather*.

"<hali>"
"hali" N 9/10-SG { the } { state , condition , essence }
"<ya>"
"ya" GEN-CON 9-SG { of }
"<hewa>"
"hewa" N 9/10-SG { the } { air , weather , atmosphere }
PLACE
"<inayobadilika>"
"badilika" V 9-SG-SP VFIN { it } PR:na 9-SG-REL { which } z
[badili] { change , substitute , adapt } SV STAT

Katika hatua mbili zifuatazo tunaona jinsi tungo *hali ya hewa* inavyobadilika na kuwa neno moja weather na muundo wa kitenzi inayobadilika inachukuliwa kama kivumishi katika muktadha huu.

- a) "<hali_ya_hewa>"
"hali_ya_hewa" N 9/10-SG { the } { weather }
"<inayobadilika>"
"badilika" V 9-SG-SP VFIN { it } PR:na 9-SG-REL { which } z
[badili] { change , substitute , adapt } SV STAT
- b) "<hali_ya_hewa>"
"hali_ya_hewa" N 9/10-SG { the } { weather }
"<inayobadilika>"
ADJ ADJ-PR-REL { change }

Tafsiri ya *change* itabadilishwa na kuwa kitenzi nomino *changing* na pia kubadilisha mpangilio vya vipengele vya tungo na kupata *The changing weather*

Data asilia kutoka kwa (Hurskainen 2014:6) Ripoti namba:15

MFANO 73

Mfano huu unaonyesha aina nyingine ya jinsi vivumishi ambavyo ni tungo-funge za Kiswahili zinazotumia umilikishi huwa zinashughulikiwa na kanuni za kisarufi kikompyuta na kutafsiriwa

kuwa kivumishi cha neno moja la Kiingereza na hata kurekebisha mpangilio wa maneno mathalan;

mtu *mwenye uwezo* = *Able man*

```
"<mtu>"
"mtu" N 1/2-SG HUM { the } { man , :human being , person }
"<mwenye>"
"enye" POSS-PRON :OTE 1-SG { with , possessor of }
"enye" POSS-PRON :OTE 18-SG { in }
"enye" POSS-PRON :OTE 1-SG { who }
"enye" POSS-PRON :OTE 18-SG { which }
"<uwezo>"
"uwezo" ADJ <MW { able }
```

MFANO 74

```
"<mtu>"
"mtu" N 1/2-SG HUM { the } { man }
"<mwenye>"
"enye" MW>
"<uwezo>"
"uwezo" ADJ <MW { able }
```

Data asilia kutoka kwa (Hurskainen 2014:6) Ripoti namba:15

Katika data iliyowasilishwa katika **MFANO 73** hadi **MFANO 74** tunaona kuwa kando na tafsiri ya neno lenyewe ambalo ni kivumishi kamili kuna hatua mbili zaidi za kukabiliana na upungufu wa vivumishi vya Kiswahili katika tafsiri kutoka Kiingereza. Hurskainen 2014a:1 anapendekeza matumizi ya njia tatu za (a) kutumia tungo za kuonyesha asili k.m. *mtu wa asili* (aboriginal man); (b) tungo za kuonyesha umiikaji k.m. *mtu mwenye uwezo* (able man); (c) kutumia miundo ya vitenzi pamoja na viambishi virejeshi k.m. *-ina_badilika*. Kwa kuwa tayari kuna hali kama hizi katika lugha ya Kiswahili, basi utata unaotokea unatatuliwa kutumia kanuni za uteuzi na sarufi kikwazo baada ya uchambuzi wa kimofolojia na hatimaye muktadha wa matumizi na kanuni za lugha pokezi kuhusishwa katika mpangilio wa maneno.

4.6.2. Tafsiri-mashine na vivumishi katika Kingereza kwenda Kiswahili

Mifano ya 75, 76 na 77 pia inaonyesha michakato hii kutumia miundo tofauti ya vivumishi ambavyo pia ni tungo-funge za Kiingereza na zinavyotafsiriwa kioto kwenda kwa Kiswahili.

MFANO 75

a) Sentensi ya Kiingereza hapa ni *-Non-toxic plants do not grow in waterless places*

```
"<Non-toxic>"
"non-toxic" { i+siyo+ na sumu } A-REL %A> A ABS CAPINIT
"<plants>"
"plant" { mi+mea } %SUBJ N NOM PL
"<do>"
```

"do" { AUX } SVO %+FAUXV V PRES SP
 "<not>"
 "not" { si } %ADVL NEG-PART
 "<grow>"
 "grow" { ha+i+ot+i } SVO %-FMAINV V INF
 "<in>"
 "in" { katika } %ADVL PREP
 "<waterless>"
 "waterless" { pa+si+po na maji } A-REL %A> A ABS
 "<places>"
 "place" { mahali } %<P N NOM PL
 "<<s>>"
 "<s>"

MFANO 76

b) Hali ya sentensi -*Non-governmental organizations are non-profit enterprises*

"<Nongovernmental>"
 "nongovernmental" MW<
 "<organizations>"
 "organization" { asasi zi+si+zo za kiserikali } %SUBJ N NOM
 PL <MW REL
 "<are>"
 "be" { ni } MONOSLB %+FMAINV V PRES
 "<nonprofit>"
 "nonprofit" { zi+sizo na faida } A-REL %A> A ABS
 "<enterprises>"
 "enterprise" { shughuli } %PCOMPL-S N NOM PL

Data asilia kutoka kwa (Hurskainen 2014e:6) Ripoti namba:14

MFANO 77

Matokeo ya tafsiri-mashine ya mfano wa 75 na 76 imetolewa hapa katika mfano 77. Kilicho dhahiri ni kuwa mchakato wa uainishaji maneno kimofolojia, kanuni za uondoaji utata na kanuni za kisarufi za kisintaksia za kupangilia upya kwa mujibu wa lugha lengwa au pokezi husaidia tafsiri-mashine.

a) Mimea isiyo na sumu haioti katika mahali pasipo na maji

b) Asasi zisizo za kiserikali ni shughuli zisizo na faida

Data asilia kutoka kwa (Hurskainen 2014e:6) Ripoti namba:14

Katika kuchunguza data iliyowasilishwa na kuchanganuliwa katika sehemu hii utaona kuwa vivumishi vya Kiswahili vinavyoundwa kutokana na tungo-funge za Kiingereza baada ya tafsiri vinakuwa na miundo ya viambishi rejeshi na hivyo hukamilisha tafsiri. Hurskainen 2014e anasema kuwa upungufu wa vivumishi katika Kiswahili si tatizo kwa tafsiri-mashine kwa hivyo lugha yoyote inaweza kutafsiriwa kwa Kiswahili. Katika kamusi yake ya SALAMA (Hurskainen 2015), ametunga na kuongeza vivumishi vya Kiswahili vifuatavyo kama vidahizo.

4.6.2. Tafsiri-mashine ya Maswali ya Kiswahili kwa Kiingereza

MFANO 77

Mifano ifuatayo unaonyesha jinsi Hurskainen,2009:5 inavyofafanua hali ya tafsiri-mashine kutoka Kiswahili kwenda Kiingereza kwa kuonyesha jinsi –kiulizi kinavyokuwa kiima; Mfano ni *Amekuja nani?*.

a) <s>
 "<*amekuja>"
 "ja" V 1/2-SG3-SP VFIN { (s)he } PERF:me [ja] { come } SV
 CAP @FMAINVintr
 "ja" V 1/2-SG3-SP VFIN NO-SP-GLOSS PERF:me [ja] { come } SV
 CAP @FMAINVintr
 "<nani>"
 "nani" INTERROG { who } @SUBJ
 "<?>"
 "? " QUESTION-MARK { ? } **CLB
 </s>
 Data asilia kutoka kwa Hurskainen,2009:5 Ripoti namba 8

Uondoaji wa utata sasa unahusu matumizi ya kanuni ya kuondoa koundoa kipashio cha kiima kwa kuwa kuna kiima fiche (@SUBJ) Matokeo yanakuwa;

b) <s>
 "<*amekuja>"
 "ja" V 1/2-SG3-SP VFIN NO-SP-GLOSS PERF:me [ja] { come } SV
 CAP @FMAINVintr
 "<nani>"
 "nani" INTERROG { who } @SUBJ
 "<?>"
 "? " QUESTION-MARK { ? } **CLB
 </s>
 Data asilia kutoka kwa Hurskainen,2009:5 Ripoti namba 8

Katika tafsiri hii, hatua ya kubadilisha mpangilio wa maneno –SVO- lazima ufanyike na hivyo kupata sentensi katika mstari mmoja ili kuweza kurahisisha uundaji wa kanuni tawala kutoka mfano huu.

c) (<s>) (V 1/2-SG3-SP VFIN NO-SP-GLOSS PERF:me { come } SV CAP)
 (INTERROG { who } @SUBJ) ("?" QUESTION-MARK { ? } **CLB) (</s>)

Hadi mfano huu ambapo swali litageuka '*Nani amekuja*' na kupata tafsiri '*Who has come?*;

d) (<s>) (INTERROG { who } @SUBJ) ({ has }) (V 1/2-SG3-SP VFIN NO-SP-GLOSS PERF:me { come } SV CAP) ("?" QUESTION-MARK { ? } **CLB) (</s>)

e) <s>
 "<*nani>"
 "nani" INTERROG { who } CAP @SUBJ
 "<amekuja>"
 "ja" V 1/2-SG3-SP VFIN NO-SP-GLOSS PERF:me [ja] { come } SV
 @FMAINVintr
 "<?>"
 "? " QUESTION-MARK { ? } **CLB
 </s>

Data asilia kutoka kwa Hurskainen,2009:6 Ripoti namba 8

Tafsiri-mashine katika mofimu ya kiulizi-je?

(a)

```
<s> "<*utakwendaje>" "kwenda" V 1/2-SG2-SP VFIN { you } FUT:ta [enda] { go } SV INTERROG
{ how } CAP @FMAINVintr "kwenda" V 1/2-SG2-SP VFIN NO-SP-GLOSS FUT:ta [enda] { go }
SV INTERROG { how } CAP @FMAINVintr "<?>" "?" QUESTION-MARK { ? } **CLB <\s>
```

(b)

```
<s> "<*utakwendaje>" "kwenda" V 1/2-SG2-SP VFIN { you } FUT:ta [enda] { go } SV
INTERROG { how } CAP @FMAINVintr "<?>" "?" QUESTION-MARK { ? } **CLB
<\s>
```

(c)

```
( <s> ) ( INTERROG { how } CAP ) ( { will } ) ( { you } V 1/2SG2-SP VFIN FUT:ta { go
} SV ) ( "?" QUESTION-MARK { ? } **CLB ) ( <\s> )
```

Data asilia kutoka kwa Hurskainen,2009:6 Ripoti namba 8

Katika tafsiri-mashine zinazohusu Kiswahili kwenda Kiingereza ni muhimu tuangalie jinsi teknolojia ya lugha inavyosaidia katika matumizi ya kanuni mahsusi za kila lugha husika. Mipangilio ya maneno katika Kiswahili huwa inaweza kubadilika ila katika Kiingereza haibadiliki(Hurskainen 2009).

SALAMA imetekeleza mchakato huu na kuibuka na mbinu zinazosuluhisha mabadiliko ya mipangilio ya maneno kutoka Kiingereza kwenda Kiswahili na kinyume chake. Katika viulizi kama nani?, yupi?, wapi?, ngapi? Yanapangwa upya kutumia mkururo wa kanuni au kanuni moja. Pia ifahamike kuwa kuwa miundo mingine ya viulizi kama viishio-je? K.m. utakwendaje? Ambacho ni mofimu ya kiulizi na hali zake hubadilika kama ilivyo katika mfano

4.6.3 Tafsiri-mashine ya Biblia ya Kiswahili kwa Kiluganda

Teknolojia ya lugha pia imeweza kusaidia tafsiri-mashine ya misahafu kama Biblia kutoka kwa Kiswahili hadi kwa lugha nyingine ya Kibantu kama Kiganda. Hurskainen, 2009 anaona kuwa kazi ya kutafsiri Biblia kwa lugha za Kiafrika inayochukua zaidi ya miaka kumi kukamilisha inaweza kuharakishwa kwa kutumia tafsiri mashine. Uchunguzi wa lugha chanzi na lugha pokezi husaidiwa na uchanganuzi wa kimofolojia wa Kiswahili kwa kuwa zana zake zipo. Tafsiri-oto hii lazima ipitie kwa Kiingereza ili kupata tafsiri na kisha kugeuza kwa Kiganda. Mchakato wote huu huwa na utata wa maana hatua kwa hatua ila unaondolewa kutumia kanuni za program za SALAMA. Tutaangalia mifano michache ya sehemu ya mradi aliouhughulikia Hurskainen, 2009 katika ripoti ya 5 ya SALAMA. Mifano hii inatolewa kutoka kwa **Mathayo 28:28-20** ambapo kila mstari huchukuliwa kuwa kipashio kimoja katika awamu zote za tafsiri kama ilivyo katika mfano wa .

MFANO 79

(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.

Data asilia kutoka kwa Hurskainen,2009:6 Ripoti namba 5

MFANO 80

Kwanza kuna utenganishaji wa maneno moja moja kwa mstari

```
'...[MAT28_18/]  
*yesu  
akaja  
kwao  
,  
akasema  
nao  
,  
akawaambia  
,  
*nimepewa  
mamlaka  
yote  
mbinguni  
na...''
```

Data asilia kutoka kwa Hurskainen 2009:6 Ripoti namba 5

MFANO 81

Uchanganuzi wa kimofolojia hufuatia na inavyoonekana kila neno lina maana kadhaa(utata wa maana) na pia visawe kwa Kiingereza huongeza utata zaidi (Hurskainen 2009); Hii ni sehemu tu ya uchambuzi huo:-

```
.."<[MAT28_18/]>" LINE-CODE3  
"<*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 }....

Data asilia kutoka kwa Hurskainen 2009:9 Ripoti namba 5

MFANO 80

Hatua hii inahusu uondoaji utata wa maana kwa kuchambua kimofolojia na kuainisha kisintaksia kijuujuu bila michoro ya miti. Kwa mfano sehemu ya kibainishio ‘the’ katika Kingereza kinapata maana potovu ila kwa kuwa lugha za Kiafrika hazina vibainishio, utata unaondolewa mara moja kama ifuatavyo;

```
"<[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 } PREFER SVOO @FMAINVintr
"<,>"
", " COMMA { , }
"<*nimepewa>"
"pewa" V 1-SG1-SP VFIN { *i } PERF:me z [pa] { give } PREFER 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
```

Data asilia kutoka kwa Hurskainen 2009:10- Ripoti namba 5

MFANO 81

Hatua hii nayo inahusu kuhamisha misamiati kama unavyoona kutoka kwa Kiingereza kwenda Kiganda na inavyoonyeshwa katika mfano huu, inabakia pamoja na ngeli yake ya nomino.

```
( "<[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)
```

Data asilia kutoka kwa Hurskainen 2009:11 Ripoti namba 5

MFANO 82

Matokeo ya tafsiri-mashine hii huwa;

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.

Data asilia kutoka kwa Hurskainen 2009:14 Ripoti namba 5

4.6.4. Tafsiri-mashine ya Nomino za pekee kwenda Kiingereza

Jambo muhimu la kumakinikia hapa ni jinsi tunavyosuluhisha utata wa kimofolojia kuhusu nomino za pekee wakati neno linalotokea mwanzo wa sentensi linachukuliwa kama nomino ya pekee. Hurskainen, 2013 anapendekeza kuwa maneno ambayo yanaweza kuwa nomino pekee yaongezewe katika msamiati ulioainishwa kimofolojia yalikiwa na lama maalum ana pia anaonelea kuwa mbinu za kiuwezekano zinachangia kuondoa utata huu maana katika tafsiri.

MFANO 83

Tukiangazia mfano wa tungo hii ya Kiswahili 'Wakati Rais Jakaya Kikwete akisubiriwa'.

"<*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
 2. "<*rais>"
 "rais" N 9/6-SG { the } { *president } MALE HUM CAP
 "*rais" N TITLE { *president } AN HUM
 3. "<*jakaya>"
 "*jakaya" PROPNAME SG { *jakaya }
 4. "<*kikwete>"
 "kweta" V INF-SBJN { to } z [kweta] { crawl along } SV CAP
 "kweta" V INF-SBJN { to } z [kweta] { have a rough :time } SV CAP
 "kweta" V SBJN 7/8-SG-SP VFIN { it } z [kweta] { crawl along } SV CAP
 "kweta" V SBJN 7/8-SG-SP VFIN { it } z [kweta] { have a rough :time } SV CAP
 5. "<akisubiriwa>"
 "subiriwa" V 1/2-SG3-SP VFIN { he } COND-IF { if/when } z [subiri] { wait } SVO PASS
 "subiriwa" V 1/2-SG3-SP VFIN { he } COND:ki z [subiri] { wait } SVO PASS
 "subiriwa" V 1/2-SG3-SP VFIN { she } COND-IF { if/when } z [subiri] { wait } SVO PASS
 "subiriwa" V 1/2-SG3-SP VFIN { she } COND:ki z [subiri] { wait } SVO PASS
 "subiriwa" V 1/2-SG3-SP VFIN { he } PR:a 7/8-SG-OBJ OBJ { it } z [subiri] { wait } SVO
 PASS
 "subiriwa" V 1/2-SG3-SP VFIN { she } PR:a 7/8-SG-OBJ OBJ { it } z [subiri] { wait } SVO
 PASS
 Data asilia kutoka kwa Hurskainen 2013:2 Ripoti namba 12

MFANO 84

Uainishaji wa PROP-MUST unapongezewa kwa uchanganuzi wa jina Kikwete huwa kanuni ya kuhakikisha neno Kikwete ni nomino pekee hata iwapo litaanza sentensi na hivyo utata wa maana utaondolewa kwa kuwa maana zilizobakia pamoja na kutambaa au kujikokota si sahihi.

"<*jakaya>"
 "*jakaya" PROPNAME SG { *jakaya }
 "<*kikwete>"
 "kweta" V INF-SBJN { to } z [kweta] { crawl along } SV CAP PROP-MUST
 "kweta" V INF-SBJN { to } z [kweta] { have a rough :time } SV CAP PROP-MUST
 "kweta" V SBJN 7/8-SG-SP VFIN { it } z [kweta] { crawl along } SV CAP PROP-MUST
 "kweta" V SBJN 7/8-SG-SP VFIN { it } z [kweta] { have a rough :time } SV CAP PROP-MUST

Data asilia kutoka kwa Hurskainen 2013:3 Ripoti namba 12

MFANO 85

Mfano huu unachunguza jinsi mfumo wa tafsiri-mashine wa SALAMA unavyochanganua majina ya watu mashuhuri hasa marais katika tafsiri-mashine. Majina kama Kibaki, Nyerere, Mkapa, Kikwete na Uhuru yana utata wa maana kwa kuwa yana fasiri tofauti katika Kiswahili

1. "<*julius>"
 "*julius" PROPNAME SG { *julius } MALE
 2. "<*nyerere>"
 "nyerere" N 9/10-SG { the } { :copper bangle , :brass bangle } CAP
 "nyerere" N 9/10-PL { the } { :copper bangle , :brass bangle } CAP
 3. "<*yakaya>"
 "*yakaya" PROPNAME SG { *yakaya }
 4. "<*kikwete>"
 "kweta" V INF-SBJN { to } z [kweta] { crawl along , have a rough :time } SV CAP
 "kweta" V SBJN 7/8-SG-SP VFIN { it } z [kweta] { crawl along , have a rough :time } SV CAP
 "kweta" V INF-SBJN { to } z [kweta] { crawl along , have a rough :time } SV CAP
 5. "<*uhuru>"
 "uhuru" N 11-SG { the } { freedom , independence , liberty } CAP
 6. "<*kenyata>"
 "*kenyata" PROPNAME SG { *kenyata } MALE
 7. "<*mwai>"

"*mwai" PROPNAME SG { *mwai }
 8. "<*kibaki>"
 "baki" V INF-SBJN { to } z [baki] { remain , stay behind , be left behind } SV CAP
 "baki" V SBJN 7/8-SG-SP VFIN { it } z [baki] { remain , stay behind , be left behind } SV CAP
 "baki" V INF-SBJN { to } z [baki] { remain , stay behind , be left behind } SV CAP
 "kibaki" N 7/8-SG { the } { rest , remains } CAP
 "baki" ADV ADV:ki 5/6-SG { the } { remainder , residue , balance } CAP
 "baki" ADV ADV:ki 9/10-SG { the } { remainder , residue and balance , left-overs } CAP
 "baki" ADV ADV:ki 9/10-PL { the } { remainder , residue and balance , left-overs } CAP
 9. "<*yoveri>"
 "*yoveri" PROPNAME SG { *yoveri } MALE
 10. "<*museweni>"
 "*museweni" PROPNAME SG { *museweni }
 11. "<*benjamin>"
 "*benjamin" PROPNAME SG { *benjamin } MALE
 12. "<*mkapa>"
 "pa" V 1/2-PL2-SP VFIN { you } NARR:ka z [pa] { give } V SVOO MONOSLB CAP
 "pa" V 18-SG-SP VFIN { there } NARR:ka z [pa] { give } V SVOO MONOSLB CAP
 13. "<*ali>"
 "*ali" PROPNAME SG { *ali } MALE

Data asilia kutoka kwa Hurskainen 2013:3 Ripoti namba 12

4.6.9 Uchanganuzi wa Data

Mifano iliyowasilishwa na kuchanganuliwa katika sehemu hii inaonyesha jinsi tafsiri-mashine yenye msingi wa kikongoo na kuzingatia lugha iliyoainishwa inavyoweza kusaidia mfumo mfuma wa kuchunguza hali za tafsiri katika Kiswahili, iwe ni kuenda kwa Kiingereza au hata kwa lugha za Kibantu kama Kiganda. Kanuni za uondoaji wa utata wa komofolojia, kisemantiki na uainishaji wa kisintaksia umeonyesha jinsi unavyosaidia mfumo wa kutafsiri. Kanuni za sarufi kikwazo pia zimeonekana kuharakisha uteuzi wa kimitambo wa tafsiri inayofaa. Uchanganuzi wa kimofolojia ndiyo hasa umethibitishwa kuwa msingi wa tafsiri-mashine ifaayo kwa kuwa mifanyiko ya kisarufi huwakilishwa kwenye uchambuzi huo ambao unasaidia kuunda kanuni za kikompyuta zinazo amrisha hatua zinazoelekea kuibuka na tafsiri bora zaidi. Kwa mtaalamu anayetafiti suala la tafsiri katika lugha ya Kiswahili, tafsiri-oto itamweka katika hali nzuri zaidi kwa kuwa anachohitaji ni kucheza na kanuni za kikompyuta na kuangalia matokeo yake kitafsiri kutoka Kiswahili au kwenda Kiswahili.

4.7 Teknolojia ya lugha katika Utafiti wa Fasihi ya Kiswahili

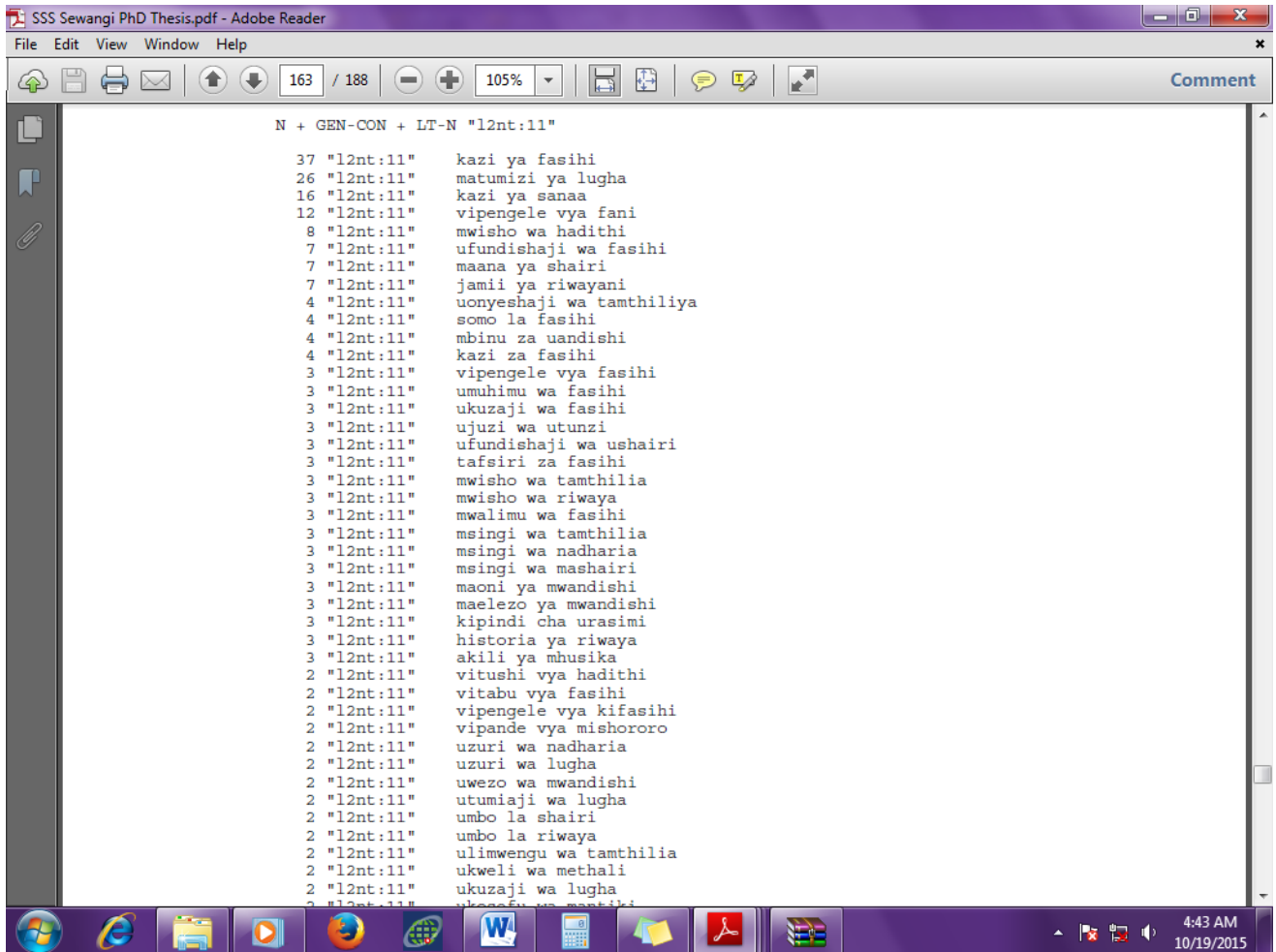
Kama tulivyodokezea katika mwanzo wa sura hii, mbinu nyingi za matumizi ya teknolojia ya lugha zinaingiliana kama jinsi vipengele vya lugha vyenyewe vinavyoingiliana. Utafiti wa kuunda kamusi kwa mfano, tumeona jinsi unavyohusika na tafsiri kwa kuwa ili neno kama kidahizo kupata maana na mfano sahihi ni lazima liwe limepata tafsiri yake ya Kiingereza ifaayo. Vile vile, utafiti wa fasihi ya Kiswahili unaweza kuonekana katika miktadha mingi tu ya shughuli za SALAMA. Kazi kuu ni ile ya Sewangi, 2001, aliyejihusisha na uchopoaji wa istilahi za nyanja maalum kama fasihi na

matibabu. Ukiangalia mfano 86 utaona msamiati alioibuka nao wa kifasihi. Kwa kutumia uchanganuzi wa kimofolojia, tafsiri, uondoaji utata wa maana na sarufi Kiswahili, kazi za kifasihi zilizo kwenye kongoo zinaweza kuchanganuliwa kwa kuchopoa mbinu za Sanaa kwa mfano tungo-funge kama nahau na methali na kuzifafanua. Mengineyo kama wahusika wanaweza kuchunguzwa kwa kuangalia ujirani wa maneno na vivumishi husika kwa mujibu wa kongoo.

MFANO 86

Text	Grammatical Form	Text	Grammatical Form
1 "l2nt:8" sanaa ya	1 "l2nt:8" mikondo ya ubunifu	1 "l2nt:8" vitendawili	1 "l2nt:8" midundo ya ngoma
1 "l2nt:8" sanaa ya ushairi	1 "l2nt:8" michezo ya	1 "l2nt:8" sanaa ya tanzu	1 "l2nt:8" washairi
1 "l2nt:8" sanaa ya nathari	1 "l2nt:8" mhusika wa	1 "l2nt:8" sanaa ya maonyesho	1 "l2nt:8" mchezoni
1 "l2nt:8" sanaa ya fasihi	1 "l2nt:8" mhakiki wa ushairi	1 "l2nt:8" riwaya za	1 "l2nt:8" mhakiki wa
1 "l2nt:8" riwaya za	1 "l2nt:8" tamthilia	1 "l2nt:8" upelelezi	1 "l2nt:8" mdundo wa ushairi
1 "l2nt:8" riwaya ya shujaa	1 "l2nt:8" mdundo wa shairi	1 "l2nt:8" riwaya ya mtunzi	1 "l2nt:8" mdundo wa muziki
1 "l2nt:8" nyimbo za taarabu	1 "l2nt:8" mawaidha ya	1 "l2nt:8" nyimbo za ngomani	1 "l2nt:8" msimulizi
1 "l2nt:8" ngoma za kijadi	1 "l2nt:8" maudhui ya	1 "l2nt:8" ngano za	1 "l2nt:8" vichekesho
1 "l2nt:8" kisimulizi	1 "l2nt:8" maudhui ya ushairi	1 "l2nt:8" ngano za	1 "l2nt:8" maudhui ya tanzu
1 "l2nt:8" fasihi simulizi	1 "l2nt:8" maudhui ya riwaya	1 "l2nt:8" nadharia ya utanzu	1 "l2nt:8" maudhui ya ngano
1 "l2nt:8" nadharia ya utanzu	1 "l2nt:8" maudhui ya	1 "l2nt:8" nadharia ya	1 "l2nt:8" kichekesho
1 "l2nt:8" urasimi	1 "l2nt:8" maudhui ya fasihi	1 "l2nt:8" nadharia ya	1 "l2nt:8" masimulizi ya
1 "l2nt:8" uhakiki	1 "l2nt:8" ngoma	1 "l2nt:8" mwimbaji wa	1 "l2nt:8" masimulizi ya
1 "l2nt:8" taarabu	1 "l2nt:8" masimulizi ya	1 "l2nt:8" mwandishi wa	1 "l2nt:8" mwandishi ya
1 "l2nt:8" tamthiliya	1 "l2nt:8" masimulizi ya	1 "l2nt:8" mwandishi wa	1 "l2nt:8" mhusika
1 "l2nt:8" tamthilia	1 "l2nt:8" hadithi	1 "l2nt:8" muundo wa	1 "l2nt:8" mashairi ya
1 "l2nt:8" muundo wa	1 "l2nt:8" masimulizi	1 "l2nt:8" usimulizi	1 "l2nt:8" masimulizi
1 "l2nt:8" muktadha wa utunzi	1 "l2nt:8" mashairi ya	1 "l2nt:8" muktadha wa utungo	1 "l2nt:8" mashairi ya jadi
1 "l2nt:8" muktadha wa shairi	1 "l2nt:8" majibizano ya	1 "l2nt:8" muktadha wa	1 "l2nt:8" kishairi
1 "l2nt:8" kitamaduni	1 "l2nt:8" mahadhi ya wimbo	1 "l2nt:8" muktadha wa	1 "l2nt:8" maandishi ya sanaa
1 "l2nt:8" iktisadi	1 "l2nt:8" maandishi ya	1 "l2nt:8" muktadha wa	1 "l2nt:8" nathari
1 "l2nt:8" hadithi	1 "l2nt:8" maandishi ya	1 "l2nt:8" mtunzi wa	1 "l2nt:8" mwandishi
1 "l2nt:8" vichekesho	1 "l2nt:8" maandishi ya	1 "l2nt:8" mtiririko	1 "l2nt:8" maandishi
1 "l2nt:8" hadithi	1 "l2nt:8" kinathari	1 "l2nt:8" hadithi	

Data asilia kutoka kwa Sewangi, 2001:155



Data asilia kutoka kwa Sewangi, 2001:167

4.8 Teknolojia ya Lugha na Ufundishaji wa Kiswahili- ICALL

Ufundishaji lugha ni eneo ambalo ni muhimu sit u kwa wataalamu wa lugha, wanafunzi, wahadhiri au wakereketwa wa lugha hiyo ila pia kwa watu wa taaluma tofauti au wanaotaka kujifundisha lugha geni. Kiswahili kama lugha inayoendelea kupata umarufu kote ulimwenguni kimekuwa kikifundishwa kutumia mbinu za kimapokeo na kuing'amuzi kwa kiasi kikubwa. Hurskainen 2010a anaripoti kuhusu mfumo wa kufundisha lugha kikompyuta ambao ukitekelezwa kwa mapana unaweza kurahisisha shughuli hii na hata kufanya ufundishaji kuwa wa kufurahikiwa na wanafunzi. Kwa lengo la utafiti huu, ufundishaji huu unasaidia watafiti wa Kiswahili kuangazia ngenzo na mikabala mbali mbali ya kuitalii lugha hii kimitambo. Mfumo huu unaojulikana kama *SWALEARN* kwingineko katika *SALAMA* na bila shaka ni wa kiithibati na kisayansi kwa kuwa unatumia kongoo na vichambuzi vya kimofolojia na programu za kuondosha utata wa maana, umetumika katika ujenzi wa hatua za ufundishaji Kiswahili. Mfumo huu unasemekana kuwa unafaa Zaidi kwa kuwa una mneso na mnepo wa kumpa mwanafunzi au anayejifundisha lugha uhuru wa kuteua na

kucheza na maneno anayopenda kujaribu mwenyewe (Hurskainen 2010a). Mifano ifuatayo inadhahirisha hali hii.

MFANO 88

Ili mfumo huu uwe sahihi na hata kutambua makosa na mapungufu mengine kama ni ya kiupatanisho wa kisarufi, mpangilio wa maneno kwenye sentensi au virai, mfumo wa kuunganisha maneno, kosa la hijai na kadhalika, huwa unahitaji kuandaliwa na mambo kadhaa ambayo baadhi ya hayo kwa mujibu wa Hurskainen 2010a ni; uainishaji wa kisarufi wa kila neno, kichambuzi cha kimofolojia cha kuruhusu zaidi ya neno moja kwa mstari, kuruhusu kuunganika kwa kila aina ya maneno hata yale yasiyo sahihi, maneno yasosahihi yasichanganuliwe, mfumo utambue makosa ya anayejifundisha kama mpangilio, uonyeshe kosa liko wapi na uruhusu na umwongoze mwanafunzi hatua kwa ya pili.(Hurskainen 2010a).

Hatua ya kwanza ni uchanganuzi wa kimofolojia wa sentensi inayofundishwa k.m –mti wangu ule umeanguka

mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+2-PL{my}angu
+DEM+11-SG{that}ule +V+1-SG2-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+2-PL{my}angu
+DEM+11-SG{that}ule +V+3-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+2-PL{my}angu
+DEM+11-SG{that}ule +V+11-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+3-SG{my}angu +DEM+3-
SG{that}ule +V+1-SG2-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+3-SG{my}angu +DEM+3-
SG{that}ule +V+3-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+3-SG{my}angu +DEM+3-
SG{that}ule +V+11-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+3-SG{my}angu
+DEM+11-SG{that}ule +V+1-SG2-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+3-SG{my}angu
+DEM+11-SG{that}ule +V+3-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+3-SG{my}angu
+DEM+11-SG{that}ule +V+11-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+11-SG{my}angu
+DEM+3-SG{that}ule +V+1-SG2-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+11-SG{my}angu
+DEM+3-SG{that}ule +V+3-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+11-SG{my}angu
+DEM+11-SG{that}ule +V+11-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+11-SG{my}angu
+DEM+11-SG{that}ule +V+3-SG-SP+PERF-ME{fall_down}anguka
mti wangu ule umeanguka +N+3/4-SG{tree}mti +POSS+11-SG{my}angu
+DEM+11-SG{that}ule +V+11-SG-SP+PERF-ME{fall_down}anguka

Data asilia kutoka kwa Hurskainen 2010:8- Ripoti Nambari 3-2010

Kanuni za kuondoa utata zinaundwa katika misingi ya miktadha mbali mbali. Miundo mbali mbali hufundishwa. Inahusisha kuondoa au kuteua sentensi au kipashio ambacho ni sahihi kutokana na uainishaji

huo. Mfano unaofuatia utaonyesha jinsi suala la udhibiti wa mpangilio wa maneno katika sentensi ya Kiswahili huwa na unavyomfaidi mwanafunzi au mtafiti wa ufundishaji lugha wa Kiswahili.

MFANO 89

Mpangilio wa maneno katika sarufi ya Kiswahili una vikwazo katika Sehemu Fulani ingawa ni huru kwa kiasi kikubwa. Vighairi hivi ni kama katika kiwakilishi kimilikishi hapa chini.

\$ wangu mtoto

Iwapo utaipatia kompyutawangu mtoto +? utapata jibu hili;

\$ wangu mtoto

Please check word order! Possessive pronoun can't be before a noun.

wangu mtoto +POSS+1-SG{my}angu +N+1/2-SG{child}mtoto

na pia

\$ mzuri mtoto

Please check word order! Adjective can't be before a noun.

mzuri mtoto +ADJ+1-SG{good}zuri +N+1/2-SG{child}mtoto

\$ wangu mtoto mzuri

Please check word order! Possessive pronoun can't be before a noun!

wangu mtoto mzuri +POSS+1-SG{my}angu +N+1/2-SG{child}mtoto +ADJ+1-SG{good}zuri

Data asilia kutoka kwa Hurskainen 2010a:8-Ripoti Nambari 3-2010

MFANO 90

Mfumo huu wa kujifunza Kiswahili kwa kompyuta kutumia teknolojia ya lugha unampa mwanafunzi uhuru wa kufanya kosa ili ajifunze. Mfano huu unaonyesha jinsi akifanya kosa la kitahajia hata kimaksudu mfumo wa SWALEARN utamrekebisha ifuatavyo;

\$ wtoto zangu

Please check that spelling is correct!

wtoto zangu +?

Kosa la pili ni la kiupatanishi wa sarufi hata kama tahajia iko sawa.

\$ watoto zangu

Please check the concordance!

watoto zangu +N+1/2-PL{child}mtoto +POSS+10-PL{my}angu

MFANO 91

Mfumo huu pia unampa mwanafunzi njia na mapendekezo mbali mbali ya kujifundishia kulingana

na kanuni husika za kisarufi na kongoo iliyoalamishwa na kuchanganuliwa.

\$ mti

OK. Combine this noun with adjective 'kubwa'!

mti +N+3/4-SG{tree}mti

\$ mti mkubwa

OK. Add to this string a possessive pronoun 'angu'!

mti mkubwa +N+3/4-SG{tree}mti +ADJ+3-SG{big}kubwa

\$ mti wangu mkubwa

OK. Add to this string a numeral!

mti wangu mkubwa +N+3/4-SG{tree}mti +POSS+3-SG{my}angu +ADJ+3-SG{big}kubwa

\$ mti wangu mkubwa mmoja

OK. Add to this string a demonstrative pronoun!

mti wangu mkubwa mmoja +N+3/4-SG{tree}mti +POSS+3-SG{my}angu +ADJ+3-SG{big}kubwa +NUM+3-SG{one}moja

\$ mti wangu mkubwa huu mmoja

OK. Add to this string the verb 'ota'!

mti wangu mkubwa huu mmoja +N+3/4-SG{tree}mti +POSS+3-SG{my}angu +ADJ+3-SG{big}kubwa +DEM+3-SG{this}huu +NUM+3-SG{one}moja

\$ mti wangu mkubwa huu mmoja unaota

OK. This is a full sentence with a long noun phrase. Now do the same exercise using plural forms.

Start by typing 'miti'!

mti wangu mkubwa huu mmoja unaota +N+3/4-SG{tree}mti +POSS+3-SG{my}angu +ADJ+3-SG{big}kubwa +DEM+3-SG{this}huu +NUM+3-SG{one}moja +V+3-SG-SP+PR-NA{grow}ota

\$ miti

OK. Combine this noun with adjective 'kubwa'!

miti +N+3/4-PL{tree}mti

\$ miti mikubwa

OK. Add to this string a possessive pronoun 'angu'!

miti mikubwa +N+3/4-PL{tree}mti +ADJ+4-PL{big}kubwa

Data asilia kutoka kwa Hurskainen 2010a:8-Ripoti Nambari 3-2010

Mfumo wa SWALEARN wa kujifundisha Kiswahili na hatua nyinginezo za mfumo wa kujifundisha lugha hii zilizoundwa na Hurskainen 2010, ni mbinu mwafaka ya kutafiti jinsi upatanisho wa kisarufi, uchambuzi wa kimofolojia, uondoaji wa utata na uchanganuzi lugha wa aina tofauti unavyosaidia ufundishaji lugha.

Tathmini ya mwalimu au mtafiti kwa mtu anayejifunza lugha itakuwa na ya kiithibati na kitarakimu kwa kuwa utaelewa fika ni neno gani linamsumbua, ameliandikaje, au ni masuala mengine kama sintaksi ya Kiswahili kwa mfano upatanisho wa sarufi au ni njeo na kadhalika.

4.9 Teknolojia ya lugha katika Utafiti wa Isimu Jamii na mabadiliko ya Kiswahili

Isimu jamii ambayo ni taaluma inayohusika na uchunguzi wa uhusiano wa lugha na jamii ni eneo ambalo linaweza kufaidika sana kiutafiti kwa kutumia teknolojia ya lugha na mitazamo ya kikongoo kwa kuwa lugha huhifadhiwa katika kongoo jinsi ilivyokusanywa katika matumizi yake halisi katika jamii husika, iwe ni kutoka kwa maandishi, mazungumzo na kadhalika.

SSS Sewangi PhD Thesis.pdf - Adobe Reader

File Edit View Window Help

141 / 188 105%

Comment

N + GEN-CON + N + PREP + HC-N "13nt:22"

1 "13nt:22" shimo la takataka kwa ajili ya uchafu
 1 "13nt:22" matatizo ya usafiri kwa wagonjwa
 1 "13nt:22" maisha ya udowezi ndani ya mwili
 1 "13nt:22" jukumu la wazazi katika afya
 1 "13nt:22" hatua za kinga dhidi ya minyoo

N + GEN-CON + HC-N + PREP + N "13nt:23"

1 "13nt:23" uvamizi wa ameba kwenye ini
 1 "13nt:23" upungufu wa damu kwa mama
 1 "13nt:23" umbo la mwili kwa jumla
 1 "13nt:23" ulinzi wa mwili dhidi ya magonjwa
 1 "13nt:23" sekta ya afya kwenye jumuiya
 1 "13nt:23" programu za chanjo katika jumuiya
 1 "13nt:23" magonjwa ya zinaa kwa ujumla
 1 "13nt:23" mabadiliko ya kimwili kwa wasichana
 1 "13nt:23" huduma za afya kwa mtoto
 1 "13nt:23" huduma za afya kwa kina mama

HC-N + GEN-CON + HC-V + GEN-CON + N "13nt:24"

1 "13nt:24" ogani ya kufyonzea ya mdudu

HC-N + GEN-CON + INF + GEN-CON + HC-N "13nt:26"

1 "13nt:26" madhara ya kudumu ya ugonjwa

N + GEN-CON + HC-V + GEN-CON + N "13nt:28"

1 "13nt:28" hatua ya kuzaliwa kwa mtoto
 1 "13nt:28" hali ya kuzuiwa kwa maendeleo

N + GEN-CON + HC-V + GEN-CON + HC-N "13nt:29"

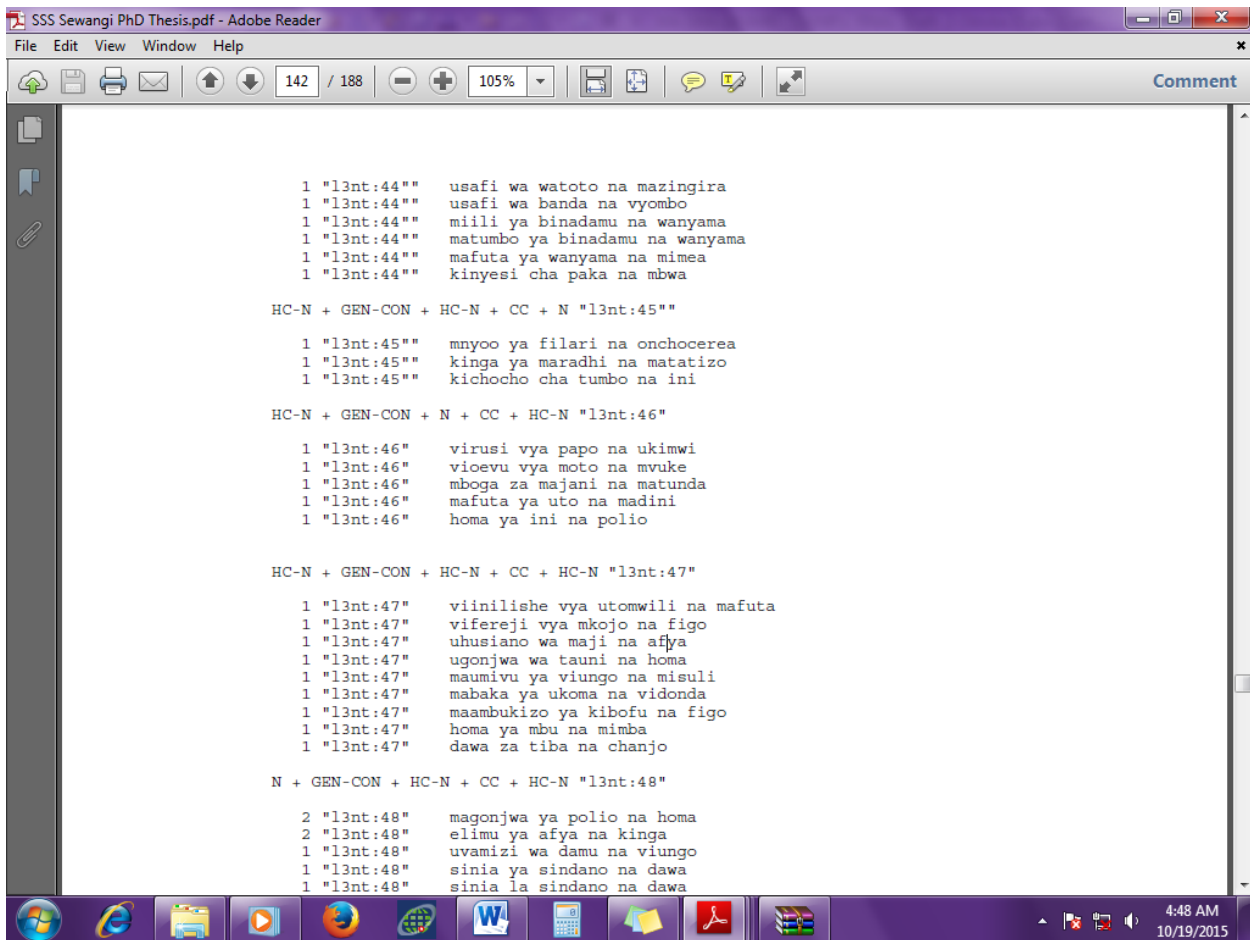
1 "13nt:29" njia ya upunguzaji wa vimelea
 1 "13nt:29" mafanikio ya kupungua kwa utapiamlo

N + GEN-CON + INF + GEN-CON + HC-N "13nt:30"

1 "13nt:30" uharibifu wa kudumu wa viungo

4:48 AM
10/19/2015

Data asilia kutoka kwa Sewangi, 2001:142



Data asilia kutoka kwa Sewangi, 2001:143

4.10 Kongoo katika Teknolojia ya lugha na Utafiti wa Kiswahili

Kipengele cha Kongoo kimeshughulikiwa kikamilifu katika sura ya tatu na kuonekana kilivyotekelezwa katika uainishaji wa vipengele vya lugha vinavyochangia kuinua utafiti wa Kiswahili katika sura hii. Mradi wa SALAMA umeonekana kuwa ni mazingira ya hali ya juu sana ya kushughulikia kongoo ya lugha na hivyo yanawezesha utekelezaji wa mfumo uliosahihi na wa kuaminika wa menejimenti ya habari kuhusu lugha ya Kiswahili (Hurskainen 2009). Katika kutafiti lugha ya Kiswahili kutumia kongoo ya HCS iliyotumika na SALAMA, uainishaji wa kimofolojia na aina nyingine ya uchanganuzi wa kisarufi hufanywa kutumia programu za kikompyuta zilizojadiliwa katika sura ya pili na kuonyeshwa katika utekelezaji wake sura ya nne.

Katika kila nyanja tumeona jinsi kongoo inavyotumika kama data muhimu ya kutafiti lugha. Ithibati kuwa mbinu zinazotumia mtazamo wa kikongoo kama ilivyo nadharia yetu, zimeibuka kidedea katika utafiti huu katika mifano kadhaa. Mmoja na wa kuaminika zaidi ni ule wa

Hurskainen, 2002 wa kutumia kongoo mbali mbalia au tuseme vijishehemu vya kongoo kutoka HCS alivyoita Korpus 1, Korpus 2 na Korpus3 katika utathmini wa kamusi zilizokuwa zinaongoza katika taaluma ya Kiswahili. Ajabu ni kuwa kamusi zote hizo zimeonekana kuwa na upungufu mkubwa, kasoro ambazo hazikuambuliwa na juhudi za wataalamu maarufu walioketi miezi na hata miaka kuunda na kuchunguza kamusi hizi.

Hii ni ithibati tosha kuwa, kongoo ni hitaji kuu katika utafiti wa lugha yoyote hasa kutumia teknolojia ya lugha. Utafiti huu uliochunguza lugha ya Kiswahili katika mazingira hayo basi, unathibitisha nadharia tete yetu ya nne kuwa kongoo ni hitaji kuu na ni muhimu katika utafiti wa lugha kikompyuta.

MFANO 94

Mfano huu unaonyesha jinsi hatua za kujaribu matini mbichi katika kongoo huwa na jinsi inavyoandaliwa kwa uchanganuzi na program mbali mbali zilizo jadiliwa kwa mujizu wa kiambatisho cha 2 ukurasa 170.

```
[1039] venus$ head /corp/swa/standard/books/books-all
<FAS> *makala za *semina ya *kimataifa ya *waandishi wa *kiswahili *iii *fasihi *taasisi
ya *uchunguzi wa *kiswahili *chuo *kikuu cha *dar_es_*salaam , 1983
<FAS> *dibaji $
<FAS> *kitabu hiki , ambacho kinatolewa katika juzuu tatu , ni matokeo ya semina mbili za
kimataifa za waandishi wa *kiswahili .$
<FAS> *semina ya kwanza ilifanyika *dar_es_*salaam , *tanzania , kuanzia tarehe 16 / 9 /
78 hadi 7 / 10 / 78 .$
<FAS> *semina ya pili vilevile ilifanyika *dar_es_*salaam kuanzia tarehe 12 / 5 / 1980 hadi
24 / 5 / 1980 .$
<FAS> *semina zote mbili zilisimamiwa na *taasisi ya *uchunguzi wa *kiswahili , *chuo
*kikuu cha *dar_es_*salaam , na kugharamiwa na *shirika la *kimataifa la *maendeleo la
*sweden ( s*i*d*a ) kupitia *shirika la *umoja wa *mataifa la *elimu *sayansi na
*utamaduni ( u*nesco )$
<FAS> *washiriki wa semina zote mbili walitoka katika nchi za *afrika *mashariki na *kati
zenye wazungumzaji wa *kiswahili , *nchi zilizoalikwa ni *burundi , *kenya , *ethiopia ,
*malagasi , *malawi , *msumbiji , *ngazija , *rwanda , *somalia , *sudan , *tanzania ,
*uganda na *zaire .$
<FAS> *zote , isipokuwa *ethiopia , *malawi , *ngazija na *zaire , zilileta wajumbe wao .$
<FAS> *kadhalika , semina zilihudhuriwa na watajamaji kutoka *sweden ( s*i*d*a ) na
u*nesco .$
<FAS> *semina zote mbili zilikuwa na shabaha za jumla zifuatazo :$
```

MFANO 95

Baada ya kuchanganuliwa neno kwa neno inapata muundo kama wa mfano huu

```
[1040] venus$ head /corp/swa/standard/books/books-all | /corp/swa/bin/one-line-
format-snt
"<<FAS>>" &
"<*makala>" "makala" N 9/10-0-PL ' article (written) ' &
"<za>" "a" GEN-CON 9/10-PL &
```

"<*semina>" "semina" N 9/10-0-SG &
"<ya>" "a" GEN-CON 9/10-SG &
"<*kimataifa>" "taifa" ADV ADV:ki 5a/6-PL 'nation. (ar)' &
"<ya>" "a" GEN-CON 9/10-SG &

Data asilia kutoka kwa www.aakkl.helsinki.fi/cameel/corpus/salamainfo.html

Hitimisho la sehemu ya Uwasilishaji wa Data

Sehemu hii imeonyesha ithibati kupitia kwa mifano ya data na maelezo yake jinsi mradi wa SALAMA ulivyochunguza lugha ya Kiswahili katika mazingira ya kikompyuta. Ni dhahiri kupitia kwa data hii kuwa utafiti wa lugha ya Kiswahili na lugha nyingine za Kiafrika kutumia mifumo changamano ya kikompyuta unawezekana na unaleta matokea bora ambayo yana manufaa zaidi ya matumizi ya utafiti wa lugha kutumia mbinu za kimapokeo. Data iliyowasilishwa hapa pamoja na ushahidi wa hojaji zilizojibiwa na mwasisi wa SALAMA na mmoja wa wanafunzi wake zinajibu maswali yetu ya utafiti na kutimiza malengo yetu sambamba na nadharia tulizoziwekea msingi kama tunavyotathmini katika. Kilichojitokeza wazi katika data hii ni kuwa utafiti wa lugha yoyote kikiwepo Kiswahili unaweza kurahisishwa kwa kuwa data inayotegemewa ipo mahali pamoja yaani kwenye kongoo-matini na inaweza kuchanganuliwa kutumia mbinu za kikompyuta ambazo zina misingi ya kuaminika kwa kufuata kanuni za kisarufi na kisayansi ambazo ni ithibati tosha ya uyakiniifu wa data hii. Natija nyingine inayopatikana kando na utegemekaji wa mbinu za kikongoo katika utafiti wa Kiswahili, ni ile istihaki ya kuwa data hiyo na mbinu hizo zinampa mtafiti nafasi pana kujaribisha kila mchakato wa kisarufi kwa muda mfupi. Isitoshe, teknolojia ya lugha ina mashiko kuliko mbinu za kimapokeo kwa kuwa ithibati ya lugha ipo katika hifadhi ya kisasa ya muundo unaosomeka kikompyuta na data hii inasasishwa jinsi muda unavyosonga na jinsi lugha inavyobadilika. Baada ya kudhihirisha kuwa haya yote yaliwezekana katika lugha ya Kiswahili kupitia kwa mradi wa SALAMA, basi pia mbinu hizi zinaweza kuwafaa wataalamu wa sasa na wajao wanaotafitia Kiswahili.

SURA YA TANO

HITIMISHO NA MAPENDEKEZO YA UTAFITI

5.0 Utangulizi

Katika sura hii tunawasilisha kwa njia pana zaidi, matokeo ya utafiti huu kuhusu nafasi ya teknolojia ya lugha katika utafiti wa lugha ya Kiswahili. Sehemu hii aidha imeshirikisha mchango wetu katika tasnia nzima ya utafiti wa lugha kutumia mbinu za kikompyuta na mustakabali wake kama ilivyodokezewa hapo awali. Katika kutathmini jinsi maswali ya utafiti huu yalivyojibika na malengo kufikiwa, bila shaka tumepima hypotheses za utafiti huu kwa misingi ya ufaafu wa nadharia mbili tulizoshirikisha.

Aidha muhtasari wa kazi hii kuhusu kila sura umetolewa. Hatimaye kumetoa mapendekezo kabambe kwa tafiti zitakazofuatia yakiwa ni pamoja na kuonyesha njia tunazofikiria kuwa zitazofaa kufuatwa ili mustakabali wa mbinu hizi za kisasa za utafiti uwe na natija kwa taaluma ya Kiswahili.

5.1 Muhtasari wa Utafiti

Utafiti huu ulilenga kubainisha mbinu na programu za kikompyuta zilizotumika katika utafiti na Uainishaji wa Lugha ya Kiswahili katika Mradi wa SALAMA na kuchunguza jinsi mbinu hizo zilivyotumika katika utafiti wa Kiswahili kwa kutumia Teknolojia ya Lugha kwenye mazingira ya huo ulioasisiwa na Prof Arvi Hurskainen katika Chuo Kikuu cha Helsinki, nchini Finland. Aidha ulikuwa na madhumuni ya kutathmini mchango wa mbinu na programu zilizotumika katika mradi huu wa SALAMA katika utafiti wa lugha ya Kiswahili kwa kutumia mbinu za teknolojia ya lugha katika siku za usoni na zinavyoweza kufaidi taaluma ya Kiswahili ha tatimaye kuchunguza na kutathmini mchango wa kongoo-matini katika utafiti wa kiteknolojia ya lugha ya Kiswahili katika Mradi wa SALAMA na unavyoweza kutumika baadaye katika utaiti.

Sura ya kwanza, imeweka msingi wa utafiti huu kwa kudurusu kazi tangulizi katika uwanja wa teknolojia ya lugha na pia tumejiwekea upeo au mipaka ya utafiti, maswali na hypotheses za utafiti huu. Aidha tumetoa sababu na msukumo wa kufanya utafiti kuhusu teknolojia ya lugha katika utafiti wa Kiswahili. Nadharia mbili zilizoongoza utafiti huu yaani Nadharia ya Msambao wa Ugunduzi na Nadharia ya Isimu Kongoo zimeelezwa kwa kina katika sura hii ikiwa ni pamoja na jinsi data za utafiti huu zitakavyokusanywa na kuchanganuliwa kwa misingi ya nadharia hizi.

Lengo la kwanza, la pili na la tatu yameshughulikiwa katika sura ya pili na ya nne wa kuonyesha jinsi mradi wa SALAMA ulivyoweza kuandaa mazingira bainifu ya kikompyuta kushughulikia

lgha ya Kiswahili. Programu mbalimbali zilizoweza kushughulikia lgha zifafanuliwa na kutathminiwa.

Programu kama SWATWOL ambazo ziliweza kuchanganua Kiswahili kimofolojia, kuondoa utata wa kimofolojia na kisemantiki, na kuweka misingi ya uainishaji wa kisintaksia pamoja na ukaguzi wa lgha kitahajia na ukaguzi wa sarufi, imeelezwa kwa upana kama uti wa mgongo wa SALAMA. Hali kadhalika kuna programu zilizohusika na uondoaji wa utata wa kisemantiki kama SWACG na pia SWAFDG inayosaidia sana katika uchambuzi wa matini za Kiswahili kisintaksia kwa kutumia mbinu za Sarufi amilifu tegemezi ya Kiswahili. Aidha kuna zana muhimu za kuunda na kutathmini kamusi za kikompyuta kama SALAMA-DC na hata za kufundisha lgha ya Kiswahili kama SWALEARN na kadhalika. Programu hizi zimebainishwa katika sura ya pili na kufafanuliwa kwa mifano na maelezo jinsi zilivyotumika katika sura ya nne.

Katika kujibu swali la nne la utafiti, kazi hii imefafanua nafasi ya hazina ya lgha iitwayo kongoo-matini ambayo ni mkusanyiko au hifadhi ya matini za lgha katika matumizi yake katika sura ya tatu. Mfano wa kongoo matini ya HCS iliyona maneno 12.5 milioni imeonyeshwa jinsi ilivyotumika kama nyenzo na pia kama mali ghafi ya kuchanganulia na kutafitia Kiswahili katika mradi wa SALAMA. Hurskainen aliongoza na bado anaendelea na ukusanyaji wa kongoo hii kutoka 1985. Nafasi ya kongoo pia imetathminiwa katika sura ya nne. Utafiti huu umeonyesha kuwa kongoo-matini inapokusanywa kutoka kwa matini za matumizi halisi ya lgha kama magazeti, vitabu, tovuti za habari au hata hotuba kama ilivyo kongoo ya HCS, basi ni kama kiwakilisho cha lgha na hivyo utafiti huwa na mashiko sana.

Utafiti hii umeonyesha kuwa matumizi ya teknolojia ya lgha kama aina fulani ya ugunduzi wa kisasa ni mbinu ambayo itabadilisha pakubwa na kuboresha viwango vya ufanisi katika utafiti wa lgha ya Kiswahili. Nadharia ya Msambao wa Ugunduzi ya Everett Rogers kwa hivyo imefaa katika mwelekeo huo. Nadharia ya Isimu Kongoo ya John Sinclair na wenzake pia imepanua na kusaidia utafiti huu wa lgha katika muktadha wa kikompyuta kwa kutumia hazina tunayoita kongoo. Nadharia zote mbinu zimedhihirisha kuwa na mashiko na kufaa zaidi uchunguzi wa lgha katika enzi ambapo teknolojia ya kompyuta imeshamiri duniani.

Imegunduliwa kuwa Kiswahili, tofauti na lgha nyingi za Kiafrika, kina hazina kubwa ya kidijitali na hivyo kinasosha kutafitiwa kwa matumizi ya teknolojia ya lgha bila changamoto zozote. Kazi hii pia imetalii uwezekano wa kufanyia nyanja zote za lgha ya Kiswahili utafiti kwa kutumia programu za kikompyuta. Maeneo kama sarufi hasa mofolojia, sintaksia, semantiki, uwakifishaji na pia nyanja za tafsiri, uundaji wa kamusi, mapisi la lgha, isimu jamii na fasihi vinaweza

kuchunguzwa kutumia mbinu hizi kwa kuwa ithibati kuwa zilifaulu katika mradi wa SALAMA inaafiki.

Hatimaye, utafiti huu umehitimisha na kutoa mapendekezo kwa tafiti zitakazofuatia kuwa juhudi zaidi zinahitajika katika uimarishaji wa mbinu za kikompyuta katika utafiti wa Kiswahili hasa ukusanyaji na uhifadhi wa kongoo matini.

5.2 Tathmini ya Maswali ya Utafiti na Upimaji wa Haipotesia

Utafiti huu uliongozwa na maswali manne. Swali la kwanza linauliza, Ni mbinu gani zinazotumika katika kufanyia utafiti lugha ya Kiswahili kutumia teknolojia ya lugha kikompyuta kwenye Mradi wa SALAMA? Tumedhihirisha kupitia kwa makala mbalimbali za mwasisi wa mradi huu zinazopakana katika tovuti ya SALAMA pamoja na ripoti za kitaalamu, majibu za hojaji za utafiti pamoja na kazi za waliuwahi kuhusika na mradi huu kuwa mbinu kadhaa zilitumika. Mbinu hizo ambazo ni programu za kikompyuta zilizojengwa kwa michakato ya kialgorithmu (kisayansi na hisabati) ni kama SWATWOL (kichanganuzi cha kimofolojia), SWACG (Kichanganuzi cha lugha kwa sarufi kikwazo), SWAFDG (Kiundosha utata wa kisemantiki na kisintaksia kwa kutumia mbinu za sarufi amilifu tegemezi), SALAMA-DC (Programu ya kuunda na kutathmini kamusi ya Kiswahili), SWALEARN (programu ya kufundishia Kiswahili kwa wageni na wanafunzi kutumia mbinu mwingiliano katika jukwaa la Kingereza), na mbinu nyinginezo ambazo zinafaulisha uondoshaji wa utata wa maana kwa kukiasi hali sahihi kama SWAGUESS, uorodheshaji (W-LIST) na kadhalika. Mbinu hizi zimefafanuliwa katika sura ya pili. Haipotesia/nadharia tete ya kwanza kwa hivyo imepata uthibitisho baada ya ukweli kujidhihiri kuwa mradi wa SALAMA ulitumia mazingira hayo katika utafiti wa Kiswahili. Kwa hivyo programu hizo zimejadiliwa na kuonyesha jinsi zinavyofaa katika utafiti wa lugha ya Kiswahili kimofolojia, kisemantiki, kisintaksia, kifasihi, kileksikografia (uundaji wa kamusi), katika tafsiri, ufundishaji wa lugha, uondoaji wa utata wa maana ya maneno na ukaguaji tahajia miongoni mwa nyingine nyingi.

Katika kutathmini swali la pili na la tatu, ambayo yalisaili ni vipi mbinu hizo tulizotaja katika swali la kwanza zilitumika katika utafiti wa Kiswahili kwa kutumia teknolojia ya lugha kwenye mazingira ya kikompyuta ya Mradi wa SALAMA na kutaka kufahamu mchango ambao zinatoa katika utafiti wa lugha ya Kiswahili kwa kutumia teknolojia ya lugha katika siku za usoni na zinaweza kufaidi vipi taaluma ya Kiswahili, kwa usanjari huo, tutaziangalia pamoja. Tumbainisha katika sura ya pili na ya nne kuwa mradi wa SALAMA umekuwa kielelezo cha uchunguzi wa Kiswahili kutumia teknolojia ya lugha kwa kuwa nyanja anuwai za lugha hii ziliweza

kushughulikiwa na matokeo yaliyowasilishwa katika utafiti huu kwa mifano na maelezo katika sura ya nne. Aidha imekuwa bayana kuwa mbinu hizi za kikompyuta ni mwelekeo wa kisasa na unaorahisisha utafiti wa lugha kwa kuwa unatumia mbinu za kisayansi na kiujarabati badala ya kutumia mbinu za kimapokeo na kiungamuzi ambazo zina shaka katika data zake. Nadharia tete ya pili iliyosema kuwa mbinu za teknolojia ya lugha zilizotumika katika utafiti wa Kiswahili kwenye mazingira ya Mradi wa SALAMA zilishughulikia lugha katika viwango mbalimbali kwa hivyo imepata uthibitisho kwa kuwa tumebainisha wazi jinsi lugha ya Kiswahili ilivyochunguzwa kimofolojia, kisemantiki, kisintaksia, kimuktadha, katika tafsiri, kileksikografia katika uundaji wa kamusi, kiuhariri katika utathmini sarufi na ukaguzi wa tahajia, kifasihi na hata katika historia ya lugha na uhifadhi wake. Nadharia tete ya tatu nayo pia inapata mashiko kwa kuwa ithibati imetolewa kuwa mbinu za teknolojia ya lugha hurahisisha baadhi ya michakato ya uchunguzaji wa lugha ya Kiswahili mathalan uundaji wa kamusi, tafsiri-oto, uondoaji utata wa kisemantiki na kadhalika. Hata hivyo ni muhimu kuangalia kuwa mahitaji ya kuandaa mazingira haya ya kikompyuta yanaweza kuwa ghali kwa kuwa zinahitajika kompyuta za kisasa za kuweka seva ya kongoo na baadhi ya programu huwa ni ghali kusanidi na kuunda. Isitoshe wataalamu wanahitaji mafunzo maalum ili waweze kutumia programu hizi katika kushughulikia Kiswahili. Licha ya hayo, mapendekezo yetu yanaelekeza kuwa shughuli zote muhimu yenye kuboresha maisha ya jamii hasa utafiti inafaa kufadhiliwa kwa kila mbinu. Fauka ya hayo imependekezwa kuwa matumizi ya programu huria za kuchunguza lugha kama Anticonc3, Linux, WordNet na kadhalika zitumike zinapofaa kwa kuwa hupatikana bila malipo.

Isitoshe, ni muhimu ikumbukwe kuwa kuna ithibati tuliyoitaja ambayo sasa inayoishi nasi kama mazao ya programu za SALAMA na hivyo kuonyesha mradi huu, sambamba na maswali ya 2 na 3 ni wa kiyakinifu na wala si wa kidhahania. SALAMA ilitumika katika kukagua tahajia ya Programu ya Kiswahili ya Microsoft Office 2007 Suite na kulainisha sentensi zake na upatanisho wa kisarufi kutumia SWATWOL mpya kipengele cha kichanganuzi cha kimofolojia, ilitumika kuunda programu za simu ya Motorola kwa Kiswahili, ilitumika kamusi ya Kiswahili-Kifini-Kiswahili kwa kutumia kongoo matini ya HCS, ilitumika katika utathmini wa kikompyuta wa kamusi 5 za Kiswahili na ripoti inapatikana katika tovuti ya SALAMA kwa Kiswahili na Kiingereza na pia ilitumika kuunda upatanisho wa kisarufi wa Biblia na Apocrypha ya Kiswahili ikiwa na muundo wenye mashina ya maneno kama maeno makuu pamoja na uainishaji wa aina za maneno kwenye upatanisho huo.

Swali la nne linauliza: Kongoo-matini ya Kiswahili inasaidiaje katika utafiti wa kiteknolojia ya lugha katika Mradi wa SALAMA sasa na katika siku za usoni? Tumebainisha wazi katika sura ya tatu, nafasi ya mkusanyiko huu katika utafiti wa lugha kikompyuta, historian a maendeleo yake aina

zake, sifa zake bainifu, kongoo ya Helsinki Corpus of Swahili, matumizi yake, manufaa yake, na pia katika sura ya nne kwa kuonyesha jinsi programu za SALAMA zilivyotumia kongoo ya HCS au kutumika kuainisha kongoo nyingine tofauti. Tukiegemea nadharia ya mtazamo wa Kongoo nafasi ya kongoo imejitokeza bayana kuwa inasaidia na itafaa katika siku za usoni katika utafiti wa lugha ya Kiswahili kwa kuwa maendeleo ya teknolojia ya kompyuta yatazidi kutilia nguvu upatikanaji wa matini kutoka kwa tovuti za Kiswahili maandishi ya vitabu na kadhalika. Licha ya kuwa haipothesia ya nne inayo sema kuwa, ‘Kongoo- matini ni hitaji kuu na kiungo muhimu sana katika utafiti unaotumia kiteknolojia ya lugha ya Kiswahili’ pia imepata uthibitisho kwa mujibu wa habari hizi, ipo haja ya kuangalia viziada na shughuli za saidizi zinazofanikisha na kuboresha uwezo wa kongoo kwa matumizi ya utafiti wa lugha. Mambo haya ni kama kualamisha kongoo, kutumia programu yenye kanuni bainifu na shirikishi ili kongoo iwe na thamani. Tumependekeza ni vipi kongoo yenye thamani kwa utafiti itajengwa.

5.3 Maoni ya Kijumla/Mchango wetu

Utafiti huu umechukua mkabala mseto. Umekuwa ni utafiti ulioshirikisha data kutoka maktabani, mtandaoni na nyanjani. Katika kuchunguza nafasi ya teknolojia ya lugha katika utafiti wa Kiswahili, tulitumia mradi wa SALAMA kama kielelezo chetu kwa kuwa ni mazingira ambayo yamekuwa mfano bora na shirikishi zaidi katika utafiti wa Kiswahili kwa miaka 30 sasa. Tumesoma kazi za Prof Arvi Hurskainen zinazopatikana katika maktaba na nyingi zilizopatikana kwenye tovuti ya SALAMA na kama makala yaliyotokea katika jarida la kisomi la Nordic Journal of African Studies. Aidha tumesoma makala na tasnifu za uzamifu za wataalamu wawili Dkt.SS Sewangi(2001) na Dkt W, Ng’ang’a (2005) ambao walishughulikia teknolojia ya Kiswahili kwa kuangalia vipengele kadhaa vya SALAMA na pia walikuwa wanafunzi wa Hurskainen, mwasisi wa SALAMA.

Utafiti huu ni wa nyanjani kwa kiasi fulani kwa kuwa ulishirikisha maswali mahsus yaliyobaini kujua jinsi programu za SALAMA zilivyofanikisha utafiti wa Kiswahili pamoja na kutaka kubainisha hali ya mradi huu sasa. Hajaji zilitumwa kwa barua pepe na kujibiwa na Hurskainen na mtaalamu mmoja. Mmoja wao hakuweza kujibu ila kazi yake ya uzamifu na makala zake zilifidia sehemu kubwa ya maswali katika hojaji hiyo tuliyoambatisha katika mwisho wa tasnifu hii.

Pia uteuzi huu wa sampuli uliongozwa na nadharia zetu mbili na maswali ya utafiti ambayo yalijibika kama ilivyolengwa.

Umekuwa ni utafiti wa kimseto katika ukusanyaji na uchanganuzi wa data ambazo zimewasilishwa

kama mifano mahsusi na maelezo bayan ya hali bainifu zilizojiri wakati wataalamu waliolengwa walitumia programu mbalimbali kutafitia Kiswahili. Matokeo hayo yamewasilishwa sambamba na malengo ya utafiti huu.

5.3.1 Ufaafu wa nadharia za utafiti

Kama tulivyodokezea mwanzoni, ni kuwa tuliteua nadharia ya Msambao wa Ugunduzi na ile ya Mtazamo wa Kiisimu Kongoo ili kutimiza malengo ya utafiti. Kama ilivyodhihirika hapo awali, teknolojia ya lugha ni mchakato wa kisasa wa kushughulikia lugha. Katika karne ya 21, masuala mengi yanaeneshwa kikompyuta na kwa mujibu ya moja ya mihimili ya Msambao wa ugunduzi ni muhimu kutembea na wakati, hasa pale kuna istihaki au manufaa kushinda mifumo iliyozeleka. Falsafa ya Rodgers 2003, ni kuwa mabadiliko yoyote ya kisasa na kiteknolojia yanayosaidia jamii, yakumbatiwe na kutekelezwa kwa upana iwezekanavyo. Tumeona jinsi programu za SALAMA zilivyofaulu kushughulikia lugha ya Kiswahili, mathalan kutathmini kamusi tano na kupata upungufu mkubwa na kufafanua hali zake mbalimbali ambazo zingechukua miaka mingi bila kuwa na hakika kama baadhi ya masuala yataibuka.

Nadharia ya pili ni ile ya Mtazamo wa Kiisimu Kongoo. Hii ni nadharia inayoibukia kuwa na umaarufu katika taaluma ya lugha kiteknolojia barani Ulaya, Asia na Marekani katika karne ya 21 ikijikita katika uchunguzi wa data ya lugha asilia iliyokusanywa katika kompyuta kwa kutumia programu mbalimbali za kikompyuta. Barani Afrika kumekuwa na juhudi kidogo sana za kutafitia lugha za Kiafrika na chache zilizopo zimekuwa ni nchini Afrika Kusini, Nigeria na hapa nchini Kenya kupitia juhudi za Peter Waiganjo, Muchemi Lawrence, Mucemi Gakuru na Ng'ang'a Wanjiku wote wakiwa wataalamu wa sayansi ya kompyuta katika Chuo kikuu cha Nairobi ambao wamejikusisha na utafiti wa Kiswahili kwa njia moja au nyingine. Kiswahili, lugha ya Kiafrika yenye hadhi kubwa zaidi barani Afrika, bado haijachangamkiwa vya kutosha katika taaluma ya teknolojia ya lugha kabla ya mradi wa SALAMA, ambao kwa kinaya zaidi hunapatikana Helsinki, Finland yaani (Ulaya). Hurskainen (2005, 2008a) anakariri maoni ya Ashford 2001, kuwa lugha hii ina hazina ya kutosha ya kidijitali na hivyo inastahili kutafitiwa kikompyuta, mtazamo unaoeleza labda kwa nini amefaulu kujenga kongoo kubwa zaidi ya Kiswahili ulimwenguni yenye maneno 12.5 milioni.

Nadharia ya kongoo kama tulivyotanguliza kuelezea, ndiyo uti wa mgongo wa teknolojia ya lugha au uhandisi lugha kama inavyotambilikana kwa kuwa bila hazina ya lugha halisi katika kusanyo la matini inayosomeka kikompyuta, hatuna 'uwanja' wa kutafitia. Tangu enzi za waasisi wa mtazamo huu kama John Sinclair (1981, 2004) ambaye anatambulikana rasmi kama '*baba-wa-isimu-kongoo*',

mtazamo huu umetufaa kwa kuwa data zote zilizowasilishwa kama mifano zilichanganuliwa kutoka kwa kongoo. Pia Sewangi 2001, anasema kuwa kongoo ndio ushahidi na ithibati ya kuaminika kuwa inayotafitiwa iko jinsi inavyosemekana kuwa na inakuwa rahisi kuishughulikia upendavyo kutumia programu anuwai za tarakilishi na kutegemea malengo ya utafiti wako. Hii ni tofauti na data ya lugha ambayo hutoka kichwani mwa mtaalamu. Nadharia ya Kiisimu Kongoo kwa hivyo imeonyesha imetufaa kuangalia jinsi mradi wa SALAMA ulivyotumia kongoo ya HCS ya maneno 12.5 milioni na vijisehemu vyake mbalimbali kufanyia utafiti wa kina wa Kiswahili.

Kwa hivyo utafiti huu umedhihirisha jinsi nadharia za kisasa kama Nadharia ya Msambao wa Ugunduzi na ile ya Msimamo wa Kiisimu Kongoo zinavyoweza kutiana nguvu katika kulitalii suala zima la Kiswahili na utafiti wake katika teknolojia ya lugha. Huu ni mchango muhimu na wa pekee katika mikabala ya kiutafiti wa taaluma ya Kiswahili.

5.3.2. Changamoto za utafiti

Licha ya kuwa tumetimiza malengo yaliyowekewa utafiti huu, bado tulikumbana na changamoto kadha. Hili ni eneo jipya la kitaaluma katika utafiti wa lugha za Kiafrika. Ukosefu wa marejeleo ya tafiti za Kiswahili na kwa Kiswahili katika eneo hili la teknolojia ya lugha au uhandisi lugha kwa hivyo ilikuwa ni changamoto kuweka misingi hasa katika udurusu wa yaliyotafitiwa hapo awali. Katika idara za Kiswahili kwenye vyuo vya Afrika mashariki, hatujakumbana na tafiti zozote za kutuelekeza katika teknolojia ya lugha. Tafiti za kukaribia mkabala wetu nizile zilizofanyika kuhusu Kiswahili katika majukwaa ya kidijitali ambao ni mtazamo wa kimsingi mno unaoangalia uwepo wa Kiswahili kwenye kompyuta tofauti na michakato changamano ya uhandisi lugha wa Kiswahili. Hata hivyo, mtazamo mpana wa suala zima ulifanikisha kuweka msingi na muktadha wa utafiti wetu.

Aidha, ilikuwa ni bora, iwapo kungekuwa na fursa ya kujaribu programu angalau moja ya programu hizi kwenye matini na kushuhudia inavyofanya kazi kwa njia ya kiutekelezi. Hili lingehitaji michakato changamano ya kutafuta idhini ya kuingilia tovuti ya chuo cha Helsinki, Finland au pengine kusafiri hadi huko, shughuli ambayo bila shaka ni ghali sana. Hata hivyo malengo yetu yalitimia kwa kuwa data ilipatikana katika hazina za SALAMA mtandaoni na kwa mawasiliano na Prof Hurskainen nilipata mifano kadhaa ya kuafika upeo wa utafiti..

Licha ya kupewa idhini rasmi na kituo cha *Center for Scientific Computing* cha Chuo Kikuu cha Helsinki kutumia kongoo ya Helsinki Corpus of Swahili (HCS), kulikuwa na changamoto za kujisajili kwenye tovuti hiyo. Hili halikuathiri matokeo ya utafiti huu kwa lengo la utafiti halikuwa kuijaribu kongoo na hivyo kufaulu kuapata ripoti 19 za Hurskainen kutoka 1992-2014, ambazo ni

makala ya kitaalam kuhusu mazingira ya SALAMA, zilinifaa kwa data yote niliyohitaji.

Mhojiwa mmoja hakujibu. Hata hivyo mawazo yake yalipatikana katika kazi zake ambazo zipo mtandaoni na kufidia pungufu hilo. Fauka ya hayo , nilifanikiwa kupata majibu ya mwasisi wa SALAMA na aliyekuwa mwanafunzi wake mmoja. Kwa hivyo hilo halikuathiri utafiti huu.

Mwisho, ni kuwa utafiti huu ulihitaji mara nyingi kuwa umeunganishwa na mtandao ili upate data muhimu kwa kuwa karibu asilimia 90 ya utafiti hii imekuwa ni mtandaoni. Hili halikuwezekana kila wakati hasa nilipoondoka kutoka maeneo ya jiji. Nilikabiliana nalo kwa kupakua data nyingi iwezekanavyo wakati ninapounganishwa na mtandao na kuihifadhi ili nitumie kwa kompyuta wakati sijaunganishwa.

5.4 Mapendekezo ya utafiti

Kama tulivyosema hapo juu ni kuwa, huu ni utafiti tunaoamini kuwa ni moja ya kazi zitakazofungua dimba katika taaluma ya teknolojia lugha au uhandisi lugha wa Kiswahili. Kwa mujibu wa matokeo ya utafiti huu, mawazo ya waliohojiwa na mantiki ya changamoto zilizoibuka katika utafiti huu na tafiti nyinginezo ziliangalia lugha kikompyuta.

Tunapendekeza kuwa vyuo vinavyofundisha Kiswahili vinafaa kuanzisha kitengo cha Teknolojia ya Lugha na Isimu-Kompyuta inayojihusisha na utafiti wa Kiswahili na lugha za Kiafrika kutumia mbinu za teknolojia ya lugha au uhandisi lugha. Wanaweza kufanya hivyo kwa ushirikiano na wataalamu wa teknolojia ya lugha wenye tajriba watakaowaonyesha njia na kuwasaidia kuunda kozi ya taaluma hii mpya na kutenga bajeti mahsusi ya shughuli hiyo. Manufaa ya mbinu hii hasa katika enzi hii ya kompyuta wanafaa kuwaongoza. Kwa mfano vyuo vya Afrika Mashariki vishirikiane na vya ng'ambo vilivyopiga hatua katika uhandisi lugha kama vile Chuo kikuu cha Helsinki, Finland, chuo kikuu cha Lancaster, Uingereza, chuo Kikuu cha UNISA, Afrika Kusini n.k. ili kupata mwongozo wa kutekeleza azma hii.

Uundaji wa Kongoo kubwa ya Kiswahili unafaa kuanzishwa ukishirikisha taasisi za Utafiti, mashirika ya kifedha na makampuni ya vyombo vya habari. Utafiti kuhusu Isimu Kongoo za Kiswahili unafaa kuandamwa na kufadhiliwa kwa kuwa ndicho kipengele kitainua thamani ya lugha hii na kuipa misingi imara ya utafiti kutumia teknolojia ya lugha. Tayari kuna juhudi za taasisi ya ACALAN chini ya AU inayounda kongoo ya maneno 100 milioni. Tunapendekeza kongoo kubwa za maneno zaidi ya nusu bilioni (500,000,000) iliyo katika viwango vya kongoo za lugha za Ulaya kama Kiingereza.

Watalaamu na watafiti wa lugha ya Kiswahili wanafaa kupata mafunzo kabambe katika matumizi ya programu za kikompyuta za kuchanganua lugha kiteknolojia hatua kwa hatua. Hili litawafaa kwa kuwa na uanuwai wa stadi za kuchanganua lugha kwa mbinu za kisasa kuliko kubakia kutumia mbinu za kimapokeo kama wanaisimu au wataalamu wa lugha tu. Pia hili litafaa iwapo wanaweza kuelewa masuala ya kikompyuta na hivyo kuweza kushiriki viwango vya uchanganuzi wa lugha hizo kwa kutumia mbinu za kikompyuta kuliko kuwatafuta wataalamu wa kialgorithmu wawasaidia na uchakataji wa data za lugha. Hili linaweza kutekelezwa na ushirikiano na idara za vyuo za kompyuta.

Iwapo utafiti wa kupima mielekeo ya wataalamu wa lugha ya Kiswahili kuhusu teknolojia ya lugha utaweza basi misingi na mikakati ya utekelezaji miradi ya sampuli hii utakuwa na mshiko zaidi. Nadharia ya Msambao wa ugunduzi inaweza kutumika katika kupima ukubalifu wa ugunduzi au teknolojia mpya hii katika lugha. Hii ni kwa sababu kuna watu wanaobakia na kuganda katika hali za kimapokeo za kijamii hata jamii husika ikitaka mabadiliko (Rogers 2003).

Nadharia za Isimu, Hisabati na za kiteknolojia zishirikishwe katika teknolojia ya lugha ili kuleta mseto wa kimikabala na ushirikishi wa utafiti wa nyanja mbalimbali za lugha.

Mashirika yenye hazina kubwa za lugha ya Kiswahili na lugha nyinginezo za Kiafrika kama vile Chama cha Wachapishaji vitabu nchini Kenya KPA kishirikiane na watafiti wa Kiswahili na taasisi za utafiti kwa kuwa maneno yanayotokana na vitabu vingi vya viada vya Kiswahili wanavyochapisha mwaka baada ya mwaka watasaidia katika uundaji wa kongoo kubwa sana ya Kiswahili. Kongoo hii itafaa katika utafiti wa Kiswahili, ufunzaji wa lugha hii na hata katika uundaji programu za kusahihisha sarufi na kukagua tahajia. Vitabu vya sarufi vyenye sheria za kuaminika kiuyakinifu pia vitaundwa kutokana na kongoo hii. Programu hizi zitakazoundwa kutokana na kongoo hii zinaweza kuwa mali ya vyama hivi na wadau wake na kusaidia kila matbaa itakayochangia katika kutoa maneno ya kujengea kongoo. Wahariri wa kampuni shirika watapata mwongozo bora wa kushughulikia lugha hii kwa njia ya kuaminika baada ya kuingia ubia wa aina hii.

Wakuza mitalaa na sera za lugha nchini na taasisi zao mathalan Taasisi ya Ukuzaji Mitalaa Nchini Kenya (KICD), wanafaa kuchangia na kuhusishwa katika ujenzi wa kongoo za lugha ili miswada wanayoipata katika Kiswahili na vitabu wanavyotathmini viwe vimejikita kwa njia hakika ya kikongoo zitumiyo teknolojia ya lugha. Vigezo vyao vya utathmini wa vitabu na miswada hii

utakuwa wa kuaminika zaidi kuliko ule wa kiungamuzi wa kuteua vitabu vya kupendekezea shule. Kwa njia hii pia wataoanisha istilahi mbalimbali za Kiswahili kwa njia rahisi na hakika.

Utafiti wa lugha ya Kiswahili unafaa kujikita zaidi katika mitazamo ya Kiisimu kongoo kwa kuwa mikabala ya kikongoo ina mashiko ya kiyakinifu, kijarabati na kisayansi ya kufanyia utafiti lugha kama ilivyodhihirika katika utafiti huu na katika mifano kadhaa ya data zilizotokana na ripoti za SALAMA na utafiti wa wataalamu mbalimbali waliotafitia mradi wa SALAMA.

Wataalamu wa lugha ya Kiswahili katika vyuo vya Afrika Mashariki wanaweza kutumia teknolojia ya lugha sio tu kujipatia hazina kubwa ya habari za Kiswahili ila pia kuhifadhai lugha hii kwa mapana na marefu. Hili linaweza kufikiwa kwa kutumia programu za kutafsiri lugha zilizojengwa na mbinu za kutumia kanuni za lugha kikompyuta kama zilizotumiwa na SALAMA. Hapa watapata hazina kubwa kutoka kwa lugha zao za mama na lugha za ng'ambo kwenda Kiswahili.

Tunapendekeza wataalamu wanaotaka kuingia kwenye tasnia hii na waliopo na wanafunzi wao wajaribu iwezekanavyo kufaidika kutokana na programu huria za kikompyuta kabla kupata uwezo na stadi kuliko kubaki vyuma kiteknolojia wakisubiri bajeti kubwa kubwa kufadhilia programu ghali za kulipiwa. Programu huria zipatikanazo sasa bila malipo ni kama Linux/Ubuntu, Anticonc3, WordNet na nyingine tulizotangulia kutaja.

Sera za lugha katika mataifa ya Afrika zipanie katika kuinua lugha zao za kiasili kimajukumu na kuibuka na mikakati ya kuzishirikisha katika majukwaa na kwenye nyenzo za kikompyuta siyo tu kuzihifadhi ila pia kuzikweza kitaaluma. Hadhi ya Kiswahili Afrika Mashariki na barani Afrika nzima imekipa nafasi ya kuendelea kukwezewa kwa kuwa tayari kimekita mizizi katika mataifa husika.

5.5 Hitimisho

Utafiti huu umechunguza matumizi ya teknolojia ya lugha katika utafiti wa Kiswahili kwa kutumia mradi wa SALAMA kama kifani cha kurejelewa. Tumeweza kubainisha kuwa mradi huu ulikuwa mazingira bainifu ya kikompyuta yaliyosheheni programu anuwai zilizoweza kuchanganua, kuchunguza na kutathmini nyanja mbalimbali za lugha ya Kiswahili.

Tunaamini kuwa mchango wa utafiti huu katika taaluma hii mpya, unaweza kutumika kwa kuwafungua macho wa watafiti wa Kiswahili kwa kuyakinisha mbinu za kufanyia utafiti wa lugha hii kwa kutumia teknolojia ya lugha. Hili linawezekana na linafaa zaidi katika ufasaha wake na usasa. Iwapo mradi wenye kongoo ya maneno 12.5 milioni tu kama SALAMA ulitumiwa kufanikisha utafiti wa kina na wa viwango tulivyojadili hapa, ni kuonyesha kuwa watafiti wa

Kiswahili, hasa walio Afrika Mashariki kilipochipukia, wana fursa kubwa ya kuunda kongoo hata ya maneno milioni kadhaa na zana za kuitafiti na kutimiza mengi zaidi. Utafiti huu ni kauli mbiu kwa wataalamu hao na pia ni daro kwao waweze kulivalia njuga suala hili na kukumbatia mbinu hizi mpya. Ithibati kuwa mkabala huu wa utafiti umeanza kukubalika na wataalamu na taasisi muhimu ni hatua tuliyodokezea ya Taasisi ya ACALAN kuanzisha mradi wa kuunda kongoo ya Kiswahili na kuwa na mipango ya kuwafundisha wanaisimu kongoo katika Kiswahili.

Tunahisi kuwa utafiti huu unaleta maarifa mapya na una mchango muhimu katika taaluma ya Kiswahili katika utafiti unaoathri maisha ya jamii lugha hii siku bada ya siku kama vile, kuunda vitabu vya kisarufi kwa shule na vyuo kutumia kongoo-matini, kuunda na kutathmini kamusi za kila taaluma na sajili, kuchunguza lugha kidiskosi, kukagua tahajia ya sarufi, kutafiti mofolojia, sintaksia na semantiki ya Kiswahili. Utafiti huu umefanya hivyo kwa kutoa taarifa, ripoti na tathmini ya SALAMA ambao ni mradi wa miaka thelathini (1985-2015) na hivyo kuotoa mwongozo huu ambao anapaswa kutoa msingi wa kuzaliwa kwa taaluma mpya ya Isimu Kompyuta na Isimu Kongoo katika idara za Kiswahili na hata kuibainisha mbinu bora zaidi za tafsiri-mashine zilizofaulu zaidi ya kongoo-sambamba kama ya Google. Mapendekezo ya utafiti huu wakitekelezwa na wadau wa Kiswahili, tutaweza kuhifadhi lugha hii ambayo ni amali ya jamii katika hazina za kikompyuta kwa umbo la kongoo-matini mbalimbali kwa utafiti yenye mashiko na kwa vizazi vijavyo.

MAREJELEO

- Allwood, J., Mfusi.M, na Hendrikse, R (Wah) (2005), *Guidelines for Developing Spoken Language Corpora*. Pretoria, 2005 Southern African Language Corpora Series (Potchefstroom) Volume 1 (Potchefstroom) Juzuu 1.
- Aronoff, Mark na Janie Rees-Miller (Wah). *The Handbook of Linguistics*. Blackwell Publishing, 2002. Blackwell Reference Online. -22 May 2015 saa kumi alasiri.
- Ashford, R. (2001) Technology and the Potential Spread of Kiswahili. In Mdee, J.S. and Mwangi, H.J.M., Makala ya kongamano la kimataifa KISWAHILI 2000 Proceedings. Dar-es-Salaam: Taasisi ya Uchunguzi wa Kiswahili, pp 144-157.
- Baker, M. na Malmkjaer, K. (2005) *Routledge Encyclopaedia of Translation Studies*. New York: Taylor and Francis E-Library.
- Baker, Mona. (2007). Corpus-based Translation Studies in the Academy. *Journal of Foreign Languages*, 5.
- Beesley, K, na Karttunen, L., (2003). *Finate State Morphology*. Series: CSLI Studies in Computational Linguistics. Stanford. CA.
- Bolshakov, I. A. na Gelbukh, A. (2004) *Computational Linguistics: Models, Resources, Applications*. Mexico: Universidad Nacional Autónoma De México
- Bosch, S.E na Pretorius, L. (2002). The significance of computational morphological analysis for Zulu lexicography, in *South African Journal of African Languages*, 2002, 22.1:11-20.
- Butler, C. (1985). *Computers in Linguistics*. Worcester: Basil Blackwell
- Clark.A, Fox.C, na Lappin, S (2010) *The Handbook of Computational Linguistics and Natural Language Processing*. London. Willey-Blackwell.
- Dalrymple, M., Karttunen, L., na Shaio, S. (1987). A Morphological Analyser using Two-Level Rules. In: Dalrymple, Kaplan na wenzake. 1987. *Tools for Morphological Analysis*. Stanford.
- Dalrymple, M., Kaplan, R. M., Karttunen, L., Koskenniemi, K., Shaio, S., and Wescoat, M., (1987). *Tools for Morphological Analysis*. Stanford: Center for the Study of Language and Information. Report No. CSLI-87-108.
- Dash, N.S. (2010) *Corpus Linguistics: A General Introduction*. CIIL, Mysore, 25th August
- De Pauw G. na De Schryver G.M. (2009) Improving the computational morphological analysis of a Swahili corpus for lexicographic purposes. *Lexikos*, 18(0).
- De Pauw, G. na G.-M. de Schryver (2008). Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes. *Lexikos* 18: 303-318.
- De Pauw, G., G.-M. de Schryver na L. Levin (Wah.).(2009). *Proceedings of the EACL 2009 Workshop on Language Technologies for African Languages*. Athens: Association for Computational Linguistics.
- De Pauw, G., de Schryver, G.-M, na P.W. Wagacha (2006). Data-Driven Part-of-Speech Tagging of

Kiswahili. *Lecture Notes in Artificial Intelligence* 4188: 197-204.

De Pauw, G., P.W. Wagacha na G.-M. de Schryver.(2009). *The SAWA Corpus: A Parallel Corpus English–Swahili*. De Pauw, G. na wenzake. (Wah.). 2009: 9-16. [SAWA corpus]

De Pauw, G., & de Schryver G.-M.(2009) African Language Technology: the Data-Driven Perspective In Lesser Used Languages and Computer Linguistics (LULCLII) - Combining efforts to foster computational support of minority languages-pp 79-96 - - Bolzano, Italy Europ <https://www.csc.fi/web/research/-/helsinki-corpus-of-swahili?inheritRedirect=true>

De Schryver G-M na Prinsloo DJ. (2000). *The Compilation of Electronic Corpora, with special reference to the African languages. Southern African Linguistics and Applied Language Studies* 18: 89–106.

De Schryver, G.-M. and D. Joffe (2009). *TshwaneDJe Kiswahili Internet Corpus*. Pretoria: TshwaneDJe HLT. [TeDJe-KIC]

De Schryver, G.-M., D. Joffe, P. Joffe and S. Hillewaert (2006). Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67-83.

De Pauw, G., de Schryver, G.-M. & Wagacha, Peter Wainjo. (2009)a. A corpus-based survey of four electronic Swahili–English bilingual dictionaries. *Lexikos*, 19, p. 340–352.

Dict. 2009. The DICT Development Group [online]. <http://dict.org> *Ergane*. 2009. A Multilingual Translation Dictionary for Windows [online]. <http://download.Travlang.com/Ergane>

Fellbaum, C. (Mh.) 1998. *WordNet: An electronic lexical database*. MIT Press.

Grgurovic, Maja, (2010). "Technology-enhanced blended language learning in an ESL class: A description of a model and an application of the Diffusion of Innovations theory" Graduate Theses and Dissertations. Paper 11667.

Järvinen, T. na Tapanainen, P. (1997). A Dependency Parser for English. Technical Reports, No. TR-1. Department of General Linguistics. University of Helsinki.

Johanson, Stig (2010) Towards a Multilingual Corpus for contrastive analysis and translation studies. SPRIKreports. Oslo.Norway. Also accessed on 20th June 2015:2.15 am at <http://www.hf.uio.no/german/sprik.html>

Gikambi, H,P (2015) Language Technology in Swahili Reseach: The Building and Application of Swahili Corpus harvested fro Swahili websites and Newspapers. Unpublished paper presented at the 28th Bayreuth Swahili Colloquium, Universitat Bayreuth-Germany.

Huskainen, A.(1992).<http://filinkedincom/pub/arvi-huskainen19/A7/27A>. February 21, 2014. saa kumi.

- 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.
- Hurskainen A. (1994). Kamusi ya Kiswahili Sanifu in test: A computer system for analyzing dictionaries and for retrieving lexical data. *Afrikanistische Arbeitspapiere* 37: 169-179.
- Hurskainen A. (1995). Computer Archives of Swahili Language and Folklore: General Description. In A. Hurskainen na S.A.K. Mlacha (Wah), *Lugha, Utamaduni na Fasihi Simulizi ya Kiswahili. Dar-es-Salaam: Taasisi ya Uchunguzi wa Kiswahili (Dar-es-Salaam) na Idara ya Taaluma za Asia na Afrika (Helsinki)*. Pp. 1-15.
- Hurskainen A. (1996). Disambiguation of morphological analysis in Bantu languages. In: *Proceedings of COLING-96*, pp. 568-573.
- Hurskainen, A. (1999). SALAMA: Swahili language manager. *Nordic Journal of African Studies*, 8(2): 139-157. Available also in: www.njas.helsinki.fi
- Hurskainen A. (2002). Tathmini ya Kamusi Tano ya Kiswahili (Computer Evaluation of Five Swahili Dictionaries). *Nordic Journal of African Studies* 11(2): 283-300. Pia inapatikana katika: www.njas.helsinki.fi ilisomwa 12-05-2015 saa tatu usiku.
- Hurskainen A. 2003. New Approaches in Corpus-Based Computational Lexicography. *Lexikos* 13 (AFRILEX-reeks/series 13: 2003): 111-132.
- Hurskainen, A. (2004a). "Swahili Language Manager: A Storehouse for Developing Multiple Computational Applications". *Nordic Journal of African Studies* 13(3): 363–397. <http://www.njas.helsinki.fi>
- Hurskainen, A. (2004b). Computational testing of five Swahili dictionaries. *Proceedings of the 20th Scandinavian Conference of Linguistics, Helsinki, 7-9.1. 2004*. <http://www.ling.helsinki.fi/kielitiede/20scl/proceedings.shtml> ilisomwa 12-05-2015 saa tatu mchana.
- Hurskainen A. (2004c). *Optimizing Disambiguation in Swahili*. In *Proceedings of COLING-04, the 20th International Conference on Computational Linguistics, Geneva 23-27.8. 2004*. Pp. 254-260.
- Hurskainen, A. (2005) *Kiswahili na hali Halisi katika Teknolojia ya Lugha*, katika Madumula, J.S na Kihore, Y.M (wah) *Toleo Maalumu la Jubilei Miaka 75 ya TUKI (1930-2005) Juzuu-68. Kiswahili- Jarida la Taasisi ya Uchunguzi wa Kiswahili*.
- Hurskainen, A. (2006). *Language Technology and Resource Development in East Africa*. University of Helsinki.
- Hurskainen, A. (2008). *Nadharia katika Teknolojia ya Lugha* katika Ogechi. N.O, Shitemi.L.N na Simala.K.I, *Nadharia katika Taaluma ya Kiswahili na Lugha za Kiafrika*. Moi University. Eldoret.
- Hurskainen, A. (2009). *Intelligent Computer-Assisted Language Learning: Implementation to Swahili in Technical Reports in Language Technology Report No 3, 2009*

<http://www.njas.helsinki.fi/salama>

Hurskainen A. (2012) Prospects of language technology in Africa, 2nd Edition of Taarifa. Finnish-Tanzania Friendship Society. (Uk 10-12)

Harks, P. (2003). *Lexicography in the Oxford Handbook of computational Linguistics*. London. Oxford University.

<http://www.macmillandictionary.com/corpus.html>

Iraki, X. N., Kiswahili Software good for Language' The People's Daily, 10th May, 2004.p.16

Kamau, S.N (2005). Matumizi ya Kiswahili katika Tarakilishi.Tasnifu ya MA (Haijachapishwa), Chuo Kikuu cha Kenyatta.

Katamba, F. and Stonham J., (2006), *Morphology*.(2nd Edition) Palgrave Macmillan. Hampshire.

Khayesi, M.K (2002). Kiswahili and Information Technology' Challenges in Public University Libraries in Kenya; katika Kiswahili, Juzuu Na.5, uk. 24-31, Dar-es-Salaam: TUKI.

Kihore, M. (2000). Kiswahili na Teknolojia ya Habari.katika Mulika Nam: 26: Uk 98-105. TUKI: Dar-es-Salaam

Lacey, M. Using a New Language in Africa to Save the Dying Ones. The Standard, 16th November, 2004, pg. 10. Nairobi: Standard Limited.

Lingsoft <http://www.lingsoft.fi> 10 Januari 2015. Saa tisa alasiri.

Lingsoft <http://www.lingsoft.fi>. 10, Januari 2015. Saa tatu asubuhi.

Ms Office in Swahili as a Lip Package.<http://www.microsoft.com/downloads/details>. Februari, 22 2015, saa tano usiku.

McEnery, T na Wilson, A. 1996. *Corpus Linguistics*. Edinburgh University Press.

Mohochi, E (2002). Teknolojia ya Mawasiliano na Nafasi ya Kiswahili Ulimwenguni katika Utafiti wa Kiswahili, I.K Simala (Mhariri). Maseno: Chuo Kikuu cha Maseno.

[Muhirwe J. \(2007\)](#) Towards Human Language Technologies for Under-resourced languages in Strengthening the Roles of ICT in Development- Special Topics in computing and ICT Research. Volome 111, Kampala, Fountain Publishers (ukurasa 123-130).

Muchemi, G. (2004) *Design of Speech database for unit -selection in Kiswahili TTS*, [Internet], Oneworld, kutoka from, <http://www.llsti.org/pubs/KiswahiliSpeech_database.pdf> [Accessed May 2015]

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.

Karttunen, L., Koskenniemi, K. na Kaplan, R. M. (1987). A compiler for two-level phonological

rules. In Dalrymple, M., Kaplan, R., Karttunen, L., Koskenniemi, K., Sjaio, S., na Wescoat, M. (eds.), *Tools for Morphological Analysis*, vol. 87 - 108 of CSLI Reports, pp. 1 - 61. Center for the Study of Language and Information, Stanford University, Palo Alto, CA.

Karttunen, L. na Beesley, K. (1992). *Two-level rule compiler*. Technical Report ISTL-92-2. Xerox Palo Alto Research Center. Palo Alto, CA.

Kiputiputi, O.M. (2011) *Kamusi Sanifu ya Kompyuta*. Dar e Salaam. Taasisi ya Uchunguzi wa Kiswahili, pp. 350-376.

Kohonen, T. (1995). *Self-Organizing Maps*. Berlin: Springer. 395 *Nordic Journal of African Studies*.

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

McEnery, T na Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice* Cambridge. Cambridge University Press.

Ng'ang'a, J. (2003) *Semantic Analysis of Kiswahili Words Using the Self Organizing Map*. *Nordic Journal of African Studies*, 12(3): 407-425. Pia yapatikana katika: www.njas.helsinki.fi

Prinsloo, D.J. na de Schryver G.-M. (2001) *Taking Dictionaries for Bantu Languages into the New Millennium with Special Reference to Kiswahili, Sepedi and Isizulu*. In Mdee, J.S. and Mwansoko, H.J.M., *Makala ya kongamano la kimataifa KISWAHILI 2000 Proceedings*. Dar-es-Salaam: Taasisi ya Uchunguzi wa Kiswahili, pp 188-215.

O'Keeffe, A. & McCarthy, M. (2007). *From Corpus to classroom: Language use and language teaching*. Cambridge: CUP.

Rass, Theresa (2010) *Corpus Linguistics: An introduction to the field and its Use in Linguistics*.

Rogers, E., M. (2003) *Diffusion of Innovation. Fifth Edition*, New York- Free Press

Sewangi, S. (2000). *Tapping the neglected resource in Kiswahili terminology: Automatic compilation of the domain-specific terms from corpus*. *Nordic Journal of African Studies* 9(2): 60-84. Available also in: www.njas.helsinki.fi/pdf-Fills/vol 9 num 21 (Januari 12, 2015, saa tatu asubuhi).

Sewangi, S. (2001). *Computer-Assisted Extraction of Terms in Specific Domains: The Case of Swahili*. Ph.D. thesis. Publications of the Institute for Asian and African Studies, 1. University of Helsinki.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J (2004) *Trust the text; Language, Corpus and Discourse*. London, Routledge-Taylor and Francis.

Strazny, Phillip (mh) (2005) *Encyclopedia of Linguistics*. Vol. A-L New York. Fitzroy Dearborn (Taylor and Francis) Ltd

Stevenson, M. na Wilks, Y. (2001). *The interaction of knowledge sources in word sense*

disambiguation. *Computational Linguistics* 27(3): 321-349.

Spencer, A. na A. M. Zwicky (eds.) 1998, *The Handbook of Morphology*, Blackwell Publishing, Oxford

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

Tapanainen, P. na T. Järvinen. 1998. Dependency Concordances. *International Journal of Lexicography* 11(3): 187-203.

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

Sahin, Ismail (2006) *Detailed Review Of Rogers' Diffusion Of Innovations Theory And Educational Technology-Related Studies Based On Rogers' Theory katika The Turkish Online Journal of Educational Technology – TOJET* April ISSN: 1303-6521 volume 5 Issue 2 Article 3 14. Chuo Kikuu Cha Iowa State.

Wilks, Y. na Stevenson, M. (1998). Word sense disambiguation using optimized combinations of knowledge sources. In: *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics*, pp. 1398-1402.

Marejeleo kuhusu “Mtazamo wa Kikongoo katika Wavuti”,

<http://webascopus.org/EAL/http://www.encyclopediaofappliedlinguistics.com/>

Computational Linguistics: <http://www.mitpressjournals.org/loi/coli>

Computer Speech and Language :<http://www.elsevier.com/wps/find/journaldescription.cws_home/622808/description

Corpora: <<http://www.eupjournals.com/journal/cor>>

Corpus Linguistics and Linguistic Theory: <<http://www.degruyter.com/journals/cllt/>>

Empirical Language Research: <<http://ejournals.org.uk/ELR>>

ICAME Journal: <<http://icame.uib.no/journal.html>>

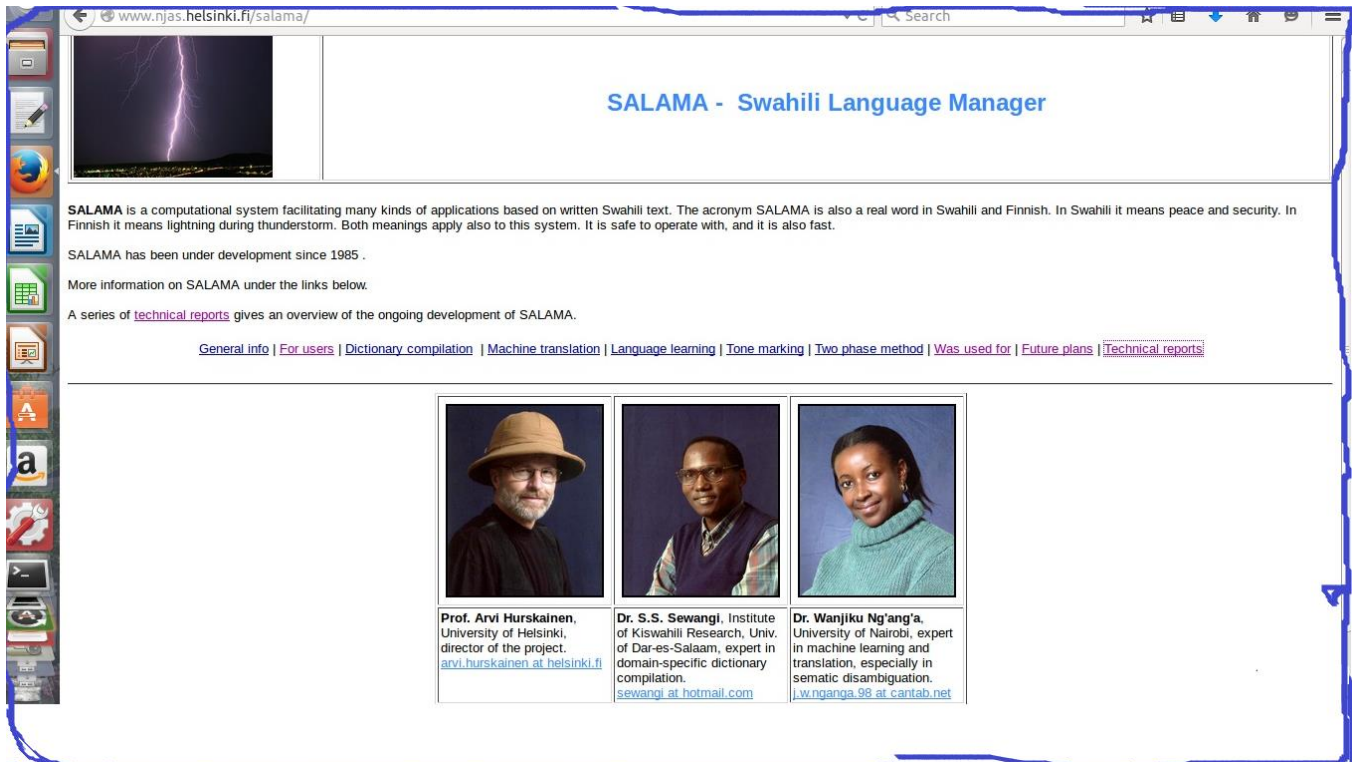
International Journal of Corpus Linguistics:<http://www.benjamins.com/cgi-bin/t_seriesview.cgi?series=IJCL>

Language Resources and Evaluation (formerly known as Computers and the Humanities):
<<http://www.springer.com/linguistics/computational+linguistics/journal/10579>>

Literary and Linguistic Computing: <<http://llc.oxfordjournals.org/>>

VIAMBATISHO

Kiambatisho 1: Ukurasa wa kwanza wa tovuti ya SALAMA (Swahili Language Manager)



www.njas.helsinki.fi/salama/

SALAMA - Swahili Language Manager




SALAMA is a computational system facilitating many kinds of applications based on written Swahili text. The acronym SALAMA is also a real word in Swahili and Finnish. In Swahili it means peace and security. In Finnish it means lightning during thunderstorm. Both meanings apply also to this system. It is safe to operate with, and it is also fast.

SALAMA has been under development since 1985 .

More information on SALAMA under the links below.

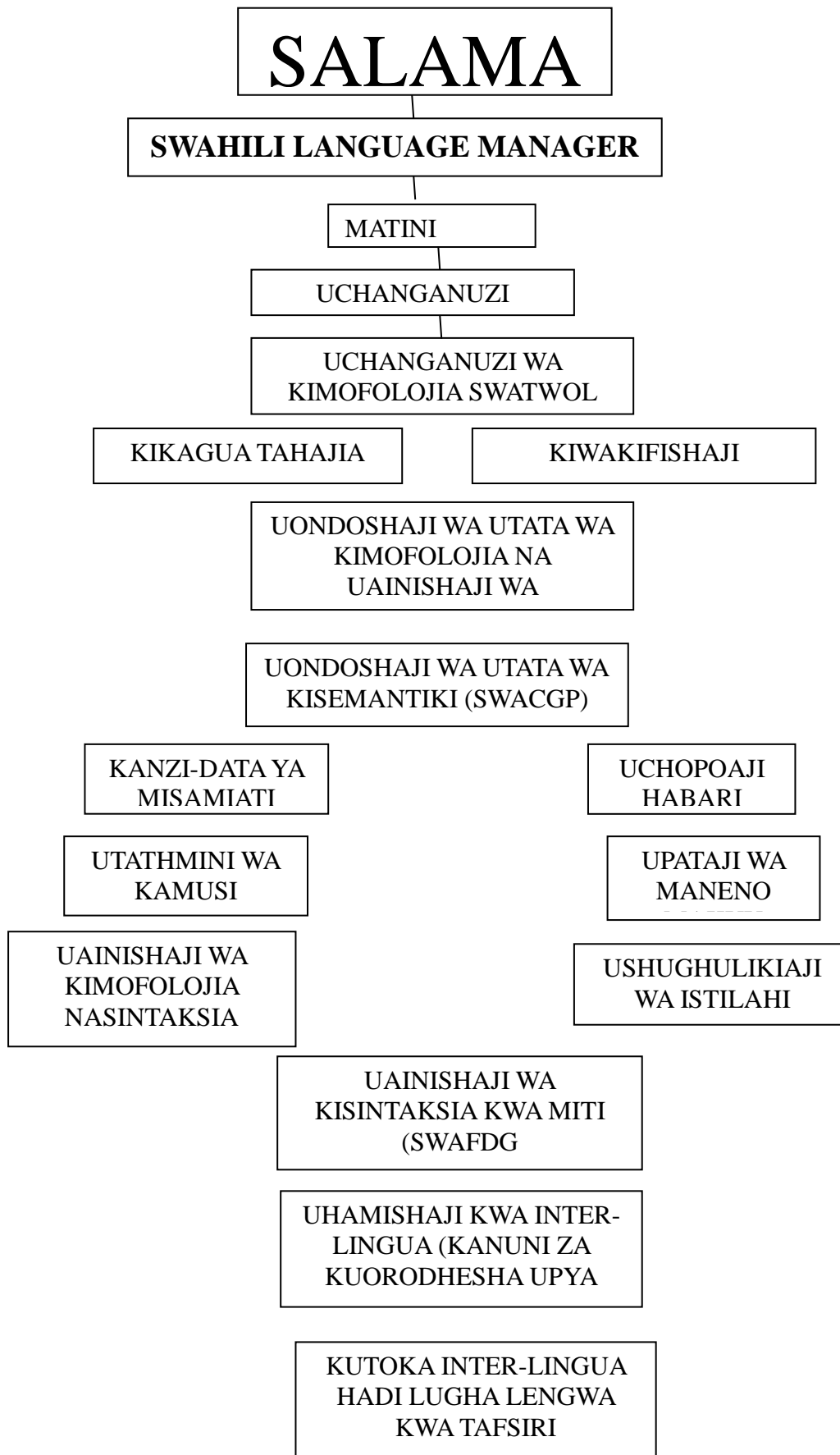
A series of [technical reports](#) gives an overview of the ongoing development of SALAMA.

[General info](#) | [For users](#) | [Dictionary compilation](#) | [Machine translation](#) | [Language learning](#) | [Tone marking](#) | [Two phase method](#) | [Was used for](#) | [Future plans](#) | [Technical reports](#)

		
<p>Prof. Arvi Hurskainen, University of Helsinki, director of the project. arvi.hurskainen@helsinki.fi</p>	<p>Dr. S.S. Sewangi, Institute of Kiswahili Research, Univ. of Dar-es-Salaam, expert in domain-specific dictionary compilation. sewangi@hotmail.com</p>	<p>Dr. Wanjiku Ng'ang'a, University of Nairobi, expert in machine learning and translation, especially in semantic disambiguation. j.w.nganga98@caritab.net</p>

Walio kwenye picha ni (kushoto) Mwasisi wa SALAMA Prof Arvi Hurskainen, (kati kati) aliyekuwa mwanafunzi wakewa uzamifu hadi 2001 na mtaalamu wa Uundaji wa Kamusi za Nyanja Mahsusni Dkt Suleiman, S.Sewangi ambaye sasa ni Katibu Mkuu wa BAKITA na (kulia) Dkt Wanjiku Ng'ang'a, pia alikuwa mwanafunzi wa uzamifu wa Prof Hurskainen hadi (2005) na mtaalamu wa Tafsiri-mashine na Ufunzaji Lugha kikompyuta hasa katika uondoshaji wa utata kisemantiki sasa ni mshauri mkuu wa teknolojia ya lugha na mhadhiri mtarazaki katika taaluma hii. Kiungo cha SALAMA ni www.njas.helsinki/salama.com/index.html

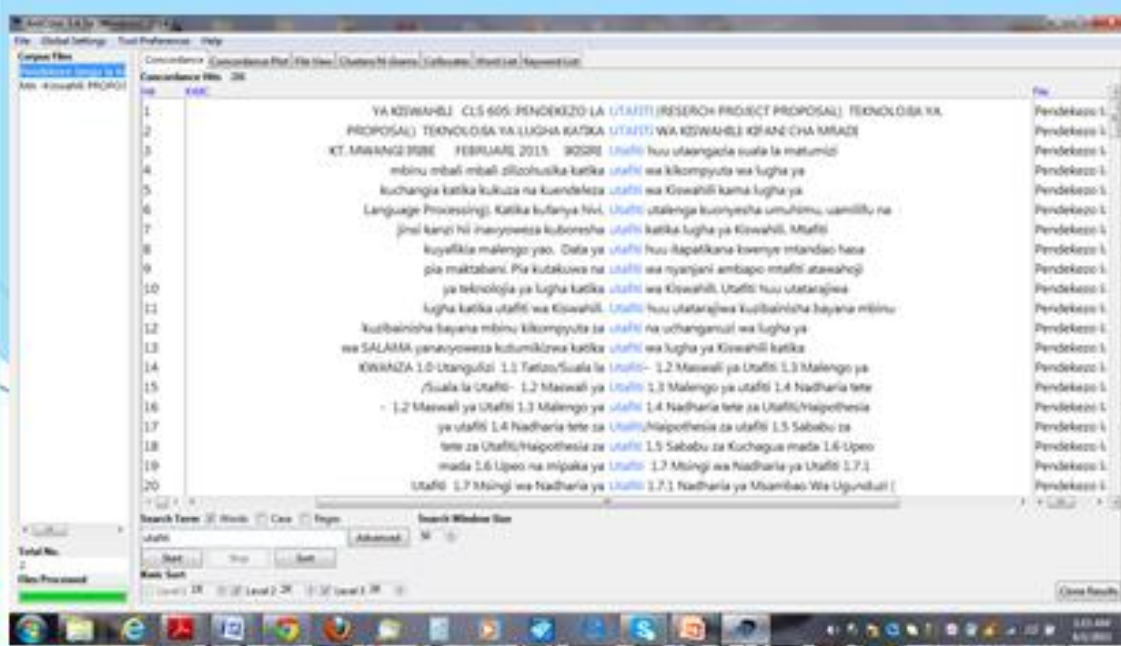
Kiambatisho 2: Jedwali linaloonyesha hatua tofauti na majukumu ya mazingira ya SALAMA



Kiambatisho 3: Matumizi ya programu huria katika utafiti wa Kiswahili kwa mitazamo ya kikongoo (Mfano wa Anticon3 na upatanisho wa sarufi wa maneno kwenye matini.

Matumizi ya Kongoo katika utafiti wa lugha ya Kiswahili

Mfano wa pendekezo langu MA- miktadha na mazingira ya neno na utokeaji wake katika mpangilio wa matini. (Concordance)



Kiambatisho 4: Sampuli ya hojaji iliyojibiwa na mwasisi wa SALAMA Prof. Arvi Hurskainen

HOJAJI KWA MWASISI WA MRADI WA SALAMA

Kwako Prof Arvi Hurskainen.

1. Umeshughulika na Mradi wa SALAMA kwa miaka mingapi sasa?

Kutoka mwaka _____ mpaka _____

2. Ni programu gani katika zile ulizotumia kuchanganulia lugha ya Kiswahili unayohisi kuwa ilikuwa na jukumu muhimu zaidi kuliko programu nyingine?

3. Je, kuna programu nyingine katika SALAMA ambazo ulizitumia na hazikutajwa au hazijaandikiwa Makala au repoti?

4. Teknolojia ya Lugha ina nafasi gani katika utafiti wa lugha ya Kiswahili kwa jumla?

5. Kwa ufupi, mradi wa SALAMA uko katika hali gani sasa? (Current status of SALAMA)

6. Ni upi mstakabali wa teknolojia ya lugha katika utafiti wa taaluma za lugha za Kiafrika kama Kiswahili? _____

-
7. Kwa maoni yako SALAMA ina changamoto zipi na inawezaje kuboreshwa zaidi?

8. Teknolojia ya lugha itasaidiaje maendeleo ya lugha za Kiafrika kama Kiswahili?

-
10. Una ushauri gani kwa taasisi za utafiti wa lugha ya Kiswahili kama vyuo vikuu kuhusu ushirikishwaje wa teknolojia ya lugha na matumizi yake katika shughuli zao?

11. Kwa maoni yako kwa nini vyuo vikuu vingi vinavyofundisha au kushughulikia lugha ya Kiswahili havijakumbatia matumizi ya teknolojia ya lugha licha ya maendeleo ya kikompyuta ya

karne ya 21?

12. Kando na zile programu za mradi wa SALAMA ni programu gani tena za kikompyuta umewahi kutumia katika kuchanganua au kutafitia lugha ya Kiswahili? Tafadhali taja na ueleze kwa ufupi kazi zake.

a) Programu _____ mwaka _____
kazi yake _____

13. Je, kongoo ya Helsinki Corpus of Swahili imeongezeka kutoka 12.5 milioni? Sasa hivi ina maneno mangapi?

14. Ulikumbana na changamoto (zozote) zipi katika ujenzi wa mazingira ya SALAMA? Unatoa tahadhari gani kwa wataalamu wa Kiswahili watumia teknolojia ya lugha wanaoibukia? (Wajiepushe na nini?)

15. Je, ni vipi kongoo kama Helsinki Corpus of Swahili inaweza kuboreshwa ili ifae zaidi katika utafiti wa lugha ya Kiswahili?

16. Je, una jambo lolote muhimu unalohisi nahitaji kulifahamu kutokana na tajriba yako ya teknolojia ya lugha ninapotafitia matumizi ya mbinu hizi katika Kiswahili?

SHUKRAN

Hezekiel Gikambi

+254722-358-448

Barua pepe: hezekieligikambi@gmail.com

Kiambatisho 5: Sampuli ya hojaji iliyopelekewa kwa wataalamu walioshiriki katika utafiti wa mradi wa SALAMA Dkt Suleiman Sewangi na Dkt Wanjiku Ng'ang'a

HOJAJI KWA WATAALAMU WA TEKNOLOJIA YA LUGHA

1. Jina -----
2. Ulishiriki katika mradi wa SALAMA kutoka mwaka_____ mpaka _____
3. Kwa ufupi tafadhali eleza jukumu lako katika SALAMA kwa muda uliobainisha hapo juu.

4. Je, ulisaidia katika ujenzi wa kipengele chochote katika mradi wa SALAMA?
Ni kipi na ulishiriki vipi?
5. Je, uliwahi kutumia mojawapo ya programu zilizoundwa katika mradi wa SALAMA katika utafiti wako?
 - a) Ndiyo
 - b) La
6. Kama jibu lako ni ndiyo je, ulitumia programu au mbinu gani ya SALAMA katika utafiti wako?
Eleza kazi yake kwa ufupi.

7. Tafadhali eleza mbinu nyingine zilizo katika SALAMA ambazo unazifahamu hata kama hukuwahi kuzitumia katika utafiti wako.

8. Kwa maoni yako programu zilizotumika katika mradi wa SALAMA zinawezaje kufaidi utafiti wa lugha ya Kiswahili?
9. Je, ni nini nafasi ya SALAMA kwa watafiti wa Kiswahili sasa hivi?

10. Kwa maoni yako SALAMA ina changamoto zipi na inawezaje kuboreshwa zaidi?

11. Teknolojia ya lugha itasaidiaje maendeleo ya lugha za Kiafrika kama Kiswahili?

12. Una ushauri gani kwa taasisi za utafiti wa lugha ya Kiswahili kama vyuo vikuu kuhusu ushirikishwaje wa teknolojia ya lugha na matumizi yake katika shughuli zao.

13. Kwa maoni yako kwa nini vyuo vikuu vingi vinavyofundisha au kushughulikia lugha ya Kiswahili havijakumbatia matumizi ya teknolojia ya lugha licha ya maendeleo ya kikompyuta ya karne ya 21?

14. Kando na zile programu za mradi wa SALAMA ni programu gani tena za kikompyuta umewahi kutumia katika kuchanganua au kutafitia lugha ya Kiswahili? Tafadhali taja na ueleze kwa ufupi kazi zake.

a) Programu _____ mwaka _____
kazi yake _____

15. Je, ni vipi wataalamu wa sasa wanaweza kuunda kongoo ya Kiswahili iwe bora kiutafiti wa lugha hii kuliko ilivyo HCS?

Kiambatisho 6: Faharasa ya alama za kimofolojia zilizotumiwa katika Tafsiri Mashine
Hurskainen(2009:1) Ripoti-8)

**CLB clause boundary

1/2-PL noun class 1/2 plural

1/2-PL3-SP noun class 1/2 third person plural subject prefix

1/2-SG noun class 1/2 singular

1/2-SG2-SP noun class 1/2 second person singular subject prefix

1/2-SG3-SP noun class 1/2 third person singular subject prefix

11/10-SG noun class 11/10 singular

3/4-SG noun class 3/4 singular

5/6-SG noun class 5/6 singular

9/10-PL noun class 9/10 plural

9/10-SG noun class 9/10 singular

<<MW multi-word expression, head two to left

<s> beginning of sentence

<\s> end of sentence

<IDIOM idiom, head one to left

<MW multi-word expression, head one to left

@<NADJ adjective modifying the noun on the left

@FMAINVintr finite main verb, intransitive

@FMAINVtr-OBJ> finite main verb, transitive but without overt object

@GCON coordinating conjunction

@SUBJ subject

A adjective

ADJ adjective

AR word of Arabic origin

A-INFL inflecting adjective

A-UNINFL uninflecting adjective

CAP word with capital initial character

CARD cardinal number

CG constraint grammar

COND:ngali conditional with the marker *ngali*

COND:nge conditional with the marker *nge*

DEM:le demonstrative with the marker *le*

FUT:ta future tense with the marker *ta*

INTERROG interrogative

MT machine translation

MW<> part of multi-word expression, other members one on the left
and one on the right

MW> part of multi-word expression, another member one on the
right

MW>> part of multi-word expression, other members one and two
on the right

MWE multi-word expression

N noun

NO-SP-GLOSS no subject prefix gloss

NOART no article

NOGLOSS no gloss

NUM numeral

NUM-INFL inflecting numeral

PAST past tense

PERF:me perfective tense with the marker *me*
POSS possessive pronoun
PR:na present tense with the marker *na*
PRON pronoun
QUESTION-MARK question mark
SG2 second person singular
STAT stative verb extension
SV subject + verb, intransitive verb
SVO subject + verb + object, transitive verb
TEMP refers to time
TIME refers to time
V verb
VFIN finite verb