Saho Corpus: Semi-automation of Verb Conjugation in Saho: Verbs Class I

Jama Musse Jama[1]
Hargeysa Cultural Centre
jama@redsea-online.org

ABSTRACT

This article develops a semi-automatic morphological analysis module *(SaCoFlexor)* to generate all inflection forms of the 585 verbs in class 1 (C-I) in the Saho language registered in the current Saho Corpus as a basic data dictionary and presents the results. *SaCoFlexor* correctly identified 98.8% of the items present in the corpus and classified them in 4 major subcategories according to the initial phonemes of the word, with the correct generation of their inflectional morphology forms generating 13,455 new words, tagged, and linked to their respective roots. The output data increased the number of words in the Saho Corpus and improved performances of the computational linguistics functions, including word frequency generation, word identification mechanism, concordance, collocations, and spell checking.

Keywords: *Saho Language, Saho Corpus, Parts-of-speech Tagging, NLP*
DOI: 10.23814/ethn.18.22.jam

1. Introduction

Saho is an East Cushitic language spoken mainly in two regions of Eritrea, in the Southern and the Northern Red Sea regions; and in the north-eastern Tigray (Ethiopia). Saho is linguistically closer to Afar language, and it is estimated that 5% of Eritreans speak Saho (Redie Bereketeab, 2010). It can be considered one of the well-described languages since a systematic scientific study of the language began in the second half of the nineteenth century with the works of Leo Reinisch in 1878 that described the Northern Saho and printed a collection of texts and a Saho (mainly Central) - German dictionary (see Reinisch, 1877, 1878a, 1878b, 1889, 1890). Prior to him, European travellers who had access to the region also documented the language with a collection of word lists published, for example, in Salt (1814) and D'Abbadie (1843) who wrote basic grammatical notes. Apart from that initial description, and because the language did not get an official orthography until recently, it remained basically spoken and had no written literature. Despite this oral prevalence, Saho in written form has been practiced since 1814 in *ad hoc* alphabets (i.e., Latin (Salt, 1841), Ge'z (Reinisch, 1878a)) and Arabic scriptures[2]), but in 1983 the Saho alphabet was officially introduced in Sahil (Eritrea) (see more on the introduction of the official alphabet Tekle Abraha, 2005; Chefena Hailemariam, 2002; and Ghirmay Negash 1999), and it was chosen a modified

[1] The author is grateful to Ahmedsaad M Omar [Axmadsacad M. Cumar] for constant support in understanding the Saho language and its grammar, as well as providing accurate corrections whenever conflicting output arrives for a specific verb conjugation calculated by the method developed by this paper. So as to Moreno Vergari and Professor Giorgio Banti.

[2] According to [Axmadsacad M. Cumar, Banti and Vergari, 2014], "It is not known when the Ajami Saho writing tradition began, because all the identified witnesses are not earlier than the last century." The oldest known Ajami script in Saho must be the work of Ibrâhîm al-Mukhtâr (1909-1969).

Latin alphabet with minor adaptation, with the first Saho books published in 1985[3]. This official orthography has 36 signs, 9 of them being digraphs, and has both long and short vowels. The Saho language has been introduced into the Eritrean national curriculum as the medium of instruction language for elementary schools in Saho-speaking regions of Eritrea after the war for dependence in 1991, with the publication of school textbooks for Science, Mathematics, Social Studies, Life Skills, and Saho Language starting 1997[4] (See Ministry of Education 2011, 2015, 2016). Different scholars described and studied the language's grammar and morphology (Ahmedsaad Mohammad Omer, Giorgio Banti, Richard Hayward, Ibrahim Mohammed Ali, Marcello Lamberti, Didier Morin, Esayas Tajebe, Moreno Vergari, Roberta Vergari, William Evert Welmers, Andrzej Zaborski, Herma Plazikowsky and Ewald Wargner, Francis Mahaffy, and others). In Banti and Vergari, 2005, where the phonology and morphology of the language have been described, the Saho verbs have been classified into four classes based on the rich inflectional system of the Saho language. The first class (C-I) are verbs with both prefix and suffix in their inflectional system. The second class (C-II) are the verbs with only suffixes in their inflectional system. The third class (C-III) are the stative verbs that indicate states and conditions. The last class (C-IV) are the compound verbs. This paper aims to limit the discussion on the first class C-I. See Banti and Vergari, 2005 & 2023 (forthcoming) for further details on the classification of the verbs and for a basic grammar of the Saho.

Saho Corpus (SaCo) is an electronic corpus of Saho language and literature that is growing fast and today contains over 250,000 tagged and classified words. The Corpus has been initiated by the author of this paper and hosted by the Hargeysa Cultural Centre[5]. With the support of linguists already working on Saho language, and the Saho-speaking community worldwide. The corpus is rapidly being populated regularly[6]. The technical structure of the corpus is derived from the online Somali Corpus (SC) (Jama Musse Jama, 2016) and it comprises a base data of 4,650 root words, with 1,201 verbs in four classes (C-I, C-II, C-III, and C-IV); 2,787 nouns; 69 particles; 39 pronouns; and other in minor numbers of parts of speech (Jama Musse Jama, 2023). Most of these entries and their classifications in Parts of Speech are coming from (Vergari and Vergari, 2003 (revised ed 2007)), but also additional root words come from other sources, including the Eritrean school curriculum textbooks (Ministry of Education, 2011, 2015, 2016), legal documents (i.e., Eritrea Constitution), other publications (collection of poetry and prose short texts), as well as social media posts, articles, and comments. The basic computational linguistics and Natural Language Processing (NLP) functionalities have been developed, and the corpus system has now *Search&Retreival*, synonym finding, word frequency lists, collocations, and concordances functions, as well as online basic spelling checker and Saho-English, Saho-Italian bilingual dictionaries. The Saho Corpus is accessible with a user interface in three languages (Saho, Somali, English) via

[3] The first books published were Grade 1 of Language, Science, and Mathematics [source Ministry of Education, Eritrea].

[4] After the first schoolbooks used during the struggle, the first new Saho schoolbooks have been published from 1997; the so-called new curriculum schoolbooks from 2004.

[5] Hargeysa Cultural Centre is a cultural hub established in 2014 in Hargeysa, Somaliland, and focusing among others research related to tangible and intangible heritage preservation.

[6] Saho Corpus was designed and directed by Jama Musse Jama (Redsea Cultural Foundation), author of the Somali Corpus Project (www.somalicorpus.com), with the collaboration of Ahmedsaad Mohammed Omer (Saahot Makaado, Ethnorêma); Giorgio Banti (University of Naples "L'Orientale", Ethnorêma), and Moreno Vergari (Ethnorêma).

www.somalicorpus.com/saho (see Jama Musse Jama, 2023, for how to use the corpus). The Corpus also contributes to East Cushitic Languages General Corpus, where the focus is on Omo-Tana Family, but the collection of data is extended to the East Cushitic Languages too (*see* Tirsit Yetbarek and Jama Musse Jama, 2022).

The verbs in class C-I, as said, are verbs with both prefixes and suffixes in their conjugation, and in the current corpus base data, they constitute 585 entries, over 10% of the total root words. The aim of this paper is to subclassify the C-1 class and find common stem generation methods for each identified subclassification and generate automatic algorithms for flexing verbs, to allow the parser of the Saho Corpus to recognize both the root words and flexed forms of the C-I verbs. To attain this target, we developed an automatic flexor module (*SaCoFlexor* V.1.0) which combines rule-based algorithms for most of the classes, and hard-coded conditions for some irregular verbs for each class identified.

Verbs are listed in the corpus in their perfect form of the first person singular as "[t]his form has been chosen as citation form because it is closer to the bare verbal stem and is easily recognized." (Banti and Vergari, 2005). C-I verbs in Saho are alliterated to the vowel *Aleph* (i.e., all verbs start with vowels: *e, i, o, u, ee, oo*. In the listed items there are no C-I verbs that start with the vowel *a* or with the long vowels *aa*, *ii* and *uu*). All the listed forms of the C-I verbs end with *e*. These tone characteristics help the application to categorize the listed verbs.

The verb *e* [to say] of which conjugations are built on the stems IMPE = *a*, SUBJV = *o*, and IMPV = *ee* is an abbreviation (spoken language) of the verb *erhxe*. Therefore, the corpus does not generate the conjugations of the verb *e* but instead, it links to *erhxe* inflectional rules.

The rest 584 verbs of the C-I class are divided into 4 major subcategories according to their initial phoneme: verbs starting with */e/, /i/, /o/* and */u/*. Each subcategory will have its own internal subclassification, for instance, differentiating the long vowel */ee/* from */e/*; and the long vowel */oo/* from */o/*. There are no C-I verbs in the data so far collected, starting with the long vowel */ii/* or */uu/*[7]. Each subclassification may have a group of verbs that does not follow the generic rule of the inflectional system of their subcategory, for instance, verbs starting with */i/* but ending with */ye/* need to be treated separately as they differ from others of the same subcategory in the Subjunctive and Imperative conjugations (see paragraph 3.1.1 for more details). Similar cases when the C-I verbs start with */u/* and end with */we/*, then they need to be treated separately as they differ from others of the same subcategory in the Subjunctive and Imperative conjugations (see paragraph 3.4.1 for more details).

2. Methodology

Apart from the perfect form of the C-I verb, the structure of SaCo database contains other three fields which construct the imperfect, subjunctive, and imperative forms for the 1st person singular of the verb respectively [Jama Musse Jama, 2023]. The contents for these fields have been collected from different sources, initially mainly from Vergari & Vergari, 2007(r), and have been extended by including resources on morphology for

[7] The current Saho Corpus database might not be exhaustive enough to confirm the absolute non-existence of such verbs. We have some (rare) Saho words starting with uu- (like *uuna*, variant of *unnay* 'be wet'). We have to remember that Esayas' grammar is only about Irob (South Saho).

those verbs that are not found from (Vergari & Vergari, 2003)[8]. *SaCoFlexor* is an automatic process of SaCo that learns from the corpus the basic behaviour of similar verbs in morphology. Its methods, whether hard-coded grammatical and morphological rules, or statistically calculated from the corpus, have been progressively developed, and with each execution time, the process 'learning' from the previous scan of the corpus. In this way, for each class of verbs, a dedicated function is written, which calculates the four stems that generate the complete inflection forms of the verb in imperfect, subjunctive, and imperative form for 1st-person, 2nd-person and 3rd-person, singular, and plural form. If the imperfect, subjunctive and imperative forms generated by the module coincide with their respective fields, the verb is declared regular in that class. Otherwise, it is listed as 'irregular', and requires a specific hard-coded programming script to generate the inflection forms. For instance, given as input the verb *eebexe* [to sell], the application *SaCoFlexor* produces the stems *eebex, aabix, aabax, ibix*, and the perfect, imperfect, subjunctive and imperative form generated from these base stems are *eebexe* [perfect form: *I sold*], *aabixe* [imperfect form: *I sell*], *aabaxo* [subjunctive form*: to sell*], *ibix* [imperative form: *sell*], which coincides with the values stored in the database, therefore *eebexe* is the regular form in this class of C-I verbs. Once declared a verb being regular, the *SaCoFlexor* application produces all inflection forms of the verb, for the case of *eebexe,* which produces the words: *eebexe, teebexe, yeebexe, neebexe, teebexin, yeebexin, aabix, taabixe, yaabixe, naabixe, taabixin, yaabixin, aabaxo, taabaxo, yaabaxo, naabaxo, yaabaxoona, ibix, ibixa, aabixii.*

The second step looks through the forms of inflection generated by the procedure on the corpus and so builds the word frequency usage for each flexed form, which gives us a confirmation if the procedures used to produce have been confirmed by the broader typology of different writings by the different authors. This second step is not interested in the correctness of the generated flexed form. However, if none of the inflection forms are used in the corpus, further research is needed to be done for the specific verb. Vice versa, if almost all flexed forms are used in different context and different documents in the corpus, this is proven that the *SaCoFlexor* is correct in producing the right inflection forms. See below the example for *erhxe* [to see]:

[8] With the automatic calculation and counterchecking with the dictionary [Vergari & Vergari, 2007(r)], a small number of spelling errors and misplacement in the dictionary has been noted and corrected. See paragraph 4.1.

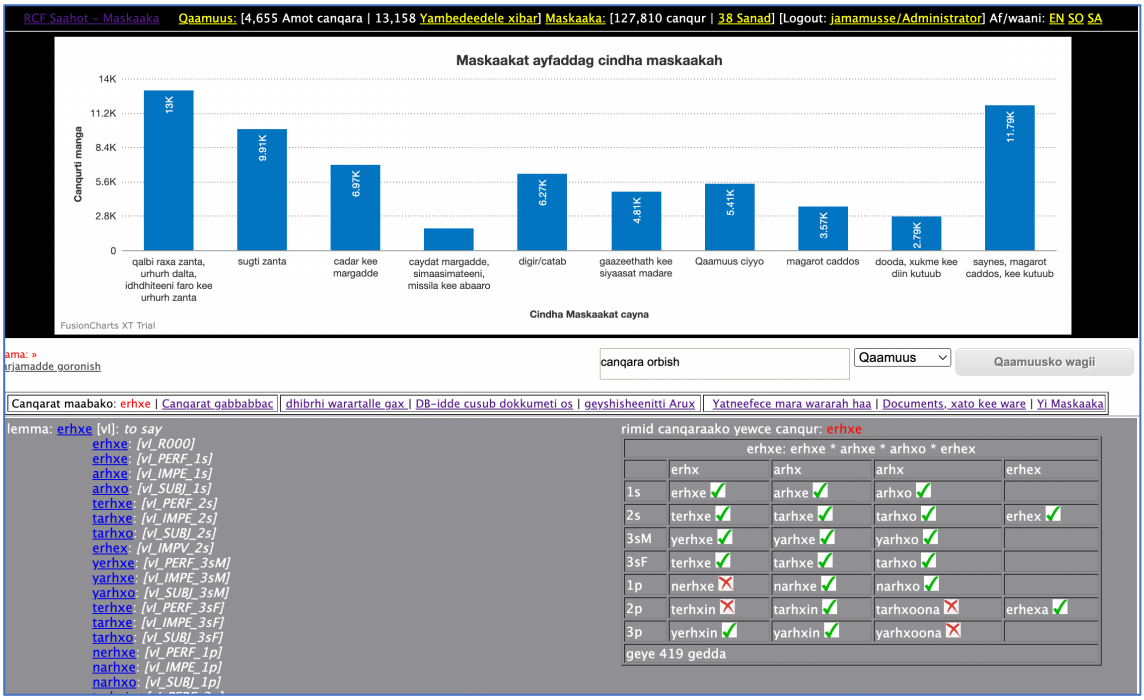

Figure 1: Screenshot of the output of the SaCoFlexor for the verb erxhe [to say].

On the left-side of the output, the tree-like structure shows the derivation form of the lemma *erhxe* of the class *vI* (in this case *vI* = = C-I verbs), the meaning of the lemma in English (this case = "to say"), and each flexed form generated correctly by the *SaCoFlexor* as well as the rule for which it was generated (i.e. *yarhxe* : [vI_IMPE_3sM = class *vI*, Imperfect form of the 3rd person singular masculine conjugation]. On the right-side table, again all generated forms are listed and their verification whether it has been used in the corpus or not. A green tick sign means that at least one time is used this flexed form in the corpus. A red cross does not mean the generated form is incorrect, but it means until now, no one used this form in the corpus. In this case the four flexed forms *nerhxe, terhxin, tarhxoona, yarhxoona* have not been used yet in the corpus when the query has been done.

Out of the 585 C-I verbs listed in the SaCo, the *SaCoFlexor* correctly produced all inflection forms for 578 verbs which is almost 99% of the base words of this verb class. In so, the module produced automatically new 12,870 correctly spelled words, tagged them for their parts of speech, and connected them to their respective root words. All modules of the SaCo (i.e. Word recognition, lemmatization, concordance generation, collocation identifier, as well as search & retrieve modules) will use these semi-automatically generated words and their relations to their respective root words (*see* Jama Musse Jama, 2023).

For example, the word *eerhege* [to know, to be familiar with], it is not found in the corpus as such, but it is present in the dictionary with all its inflectional morphology entries produced by *SaCoFlexor*. Therefore, if you search from the corpus the word *eerhege*, you will find all inflected forms of the word in the concordance two boxes as shown below.

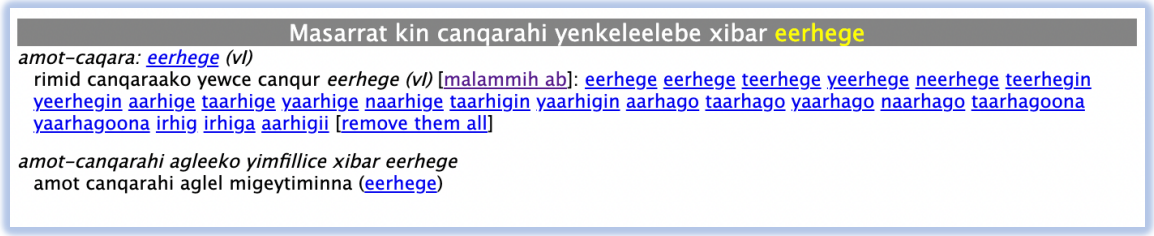

Figure 2: Screenshot from the SahoCorpus showing the current relations between the root word eerhege and its conjugations.

canqurti hoor: amot-canqarahi nafaaca eerhege

canqurti hoor amot-caqara: eerhege (vl)

TIR	Yoo mece mablo kinniki	taarhige	mara teellemko akah deecim aken.
TIR		Taarhige	maassene hido angar cadar uktub.
TIR		Aarhago	aththawwaro
TIR	madrasah yedeyinih sinni cindhamhi gondo	yaarhagoona	eltane.
TIR	tane Kaa xifzeh ayim abneh	naarhige	
TIR	Carkobkobayi tubleh	taarhige	
TIR		Taarhige	dalta ko wuli tiya tuktubeh fasleh
TIR	Xawaz xagarahi zacabah ayim	naarhige	
TIR		Taarhige	mece zantatko tiya uktub.
TIR	Jawaab tuktubeh	taarhige	Aymih
TIR	Bacshaara cadaaga geyo kee cadaaga	yaarhago	darasatti irhga geyo kee irhgo akah
TIR		Yeerhegin	zantat/qisas kataaba kee afah yashraxoona dhiican.
TIR	yexeesebe sixde yifhimiini Yifhimiini aaginah	naarhago	dhicna
TIR	Thiliyaan istimaarko	taarhige	zanta taariikh uktub.
TIR	Ishiwakalille	taarhige	baheyimhi zacabah faqhara.
TIR	ayidhdha yiksibem kee ayidhdha yikhsirem	yaarhige.	
TIR	ayidhdha yiksibem kee ayidha yikksirem	yaarhago	ciloh max-yeenimhi ishi nashaatha yaktube.
TIR	kee nifh yamxaarab mirhiicam adde	yeerhege	waqhte kinni.
TIR	Cafur cafur aseena	yaarhige.	
TIR	kee yamudhdhe gedda le xuqhuuqh	yaarhige	
TIR	dhawrhodde asheereki fadhdhitimta zuruuf	yaarhago	waajib leya.
FB	tardawo waqhte kok beyto kinam	irhig.	Ayeh Mango dhaacimtok lakalle akah abte
FB	Atu	taarhige	xiyawti yi nangalo kinni
TIR	Muluu xisbabti cayna barhisheenih	yaarhagoona	kee xaajuj araxah yashraxoona alle yatnaafacoona
TIR	Yaqhriyeti	yaarhige	
TIR	kee masarrat kin qawaacidti nizaam	yaarhagoona	kinon.
poe	Anu	aarhige	sheethan labcaadeya yoh yashshe labcaade waye
TIR	abittot kee taarikh lel lemille	yaarhagoona	kee sittiya esseroona caashir geyan.
TIR	ta margadi aamiizan luuk yinem	naarhige.	
TIR	maxaarhi abbubuh tine xiyawko garo	yeerhegin	gedda lel cadaaren yan.

Figure 3: Screenshot from the SahoCorpus showing the concordance list of the root word.

3. Classification of C-I verbs in subcategories

3.1 Subcategory 1: verbs starting with vowel *i*.

In the current Saho Corpus data dictionary, there are no C-I verbs starting with long vowel */ii/*[9], and we have 276 verbs in this subcategory, all of them following one generic rule with a minor exception of the 25 verbs ending with *"ye"* (indicated below with *) and one verb with less than 3 phonemes (indicated below with **). These have a slight

[9] see note 7

variation of inflectional system to generate the stems for Subjunctive and Imperative conjugations (see paragraph 3.1.1 and 3.1.2).

i-iwwije, ibbirhe, ibhige, ibiddile, ibissire, ibittine, ibkhile, ibrhiye, ibrike, icbide, icdire, icikkine, icimmize, icizzibe, icizzime, iclishe, icrhiye*, ictite, idbile, iddirire, ifdibe, ifdiye*, ifhime, ifiddine, ifinfine, ifittishe, ifqire, ifride, ifrime, ifrixe, ifriye*, iftixe, igdife, igdile, iggidile, igizzire, igribe, igrice, igziye*, ihjime, ikbire, ikfile, ikhzine, ikkitiye*, iklibe, iklithe, ikmile, ikshife, iksibe, iksise, ikxine, ilike, ilillige, ilimmitse, iliqhliqhe, illikkixe, imbiddile, imbissire, imbittine, imcikkine, imcizzibe, imcizzime, imfice, imfiddige, imfiddine, imfilliye*, imfiride, imfittishe, imfittixe, imfiye*, imibbixe, imirrishe, imishshithe, imjillide, imjinnine, imlike, imlillige, immicimmize, immillike, immingide, immirrixe, immissile, imniccibe, imniffiye*, imnitstsige, imsiccire, imsinniqhe, imthinqiqhe, imtikkile, imtixine, imtsiqqithe, imwiddixe, imwirrishe, imxibbishe, imxillike, imxinkishe, imzikkire, incibe, indhibbirhe, indhibbishe, indhiggile, indibbile, indime, ine**, ingiddife, ingide, inkillibe, inkillithe, inkissise, inqirribe, inqishshishe, inqissibe, inqitstsice, inqiyyide, inqiyyire, insire, insixe, inthirrife, intsige, intsiqhe, inzibbite, iqhdibe, iqhirrire, iqhishshishe, iqhiyyide, iqhliye*, iqhriye*, iqhtsice, irgide, irhibbishe, irhiggile, irhxine, irkibe, irqiqhe, isbire, iscibide, iscilishe, iscillime, iscimire, iscire, iscirhiye*, isfire, isgidife, isgigice, isgiribe, ishhide, ishiqqithe, ishrixe, ishshibixe, ishshimine, isiile, isimmire, isirrite, iskibire, iskifile, iskilithe, iskimile, iskire, iskisibe, iskitiye*, iskixine, iskiyyide, islixe, isnibe, isniqhe, isqidibe, isqiliye*, isqiriye*, isqirribe, isqithe, isqiyyire, isriye*, issifiddige, issilliye*, issiximmide, isticmire, istilime, istintine, istishhide, isxibbire, isxibire, isxibishe, isxinkishe, isxiqhiqhe, isxiriqhe, isxirrike, isxizine, itbirrice, itcibbire, itcibe, itcijjibe, itcillime, ithbice, ithhire, ithimmite, ithinqiqhe, ithirrife, ithlime, ithmice, ithrice, iththiwwire, itkibbire, itkile, itrire, itsqitsice, itsriye*, itstsibbire, itsxiye*, itticikkine, ittiminniye*, itxirrike, iwilwile, iwrishe, ixbire, ixbishe, ixfize, ixiddide, ixiggire, ixinkishe, ixkise, ixlike, ixqiqhe, ixriqhe, ixtifile, ixtime, ixzine, iybiddile, iybirike, iybirkite, iydirire, iyfiddige, iyfiddine, iyfidibe, iyfidiye*, iyfiliye*, iyfiride, iyfirime, iyfirixe, iyfiriye*, iyfissire, iyfitixe, iyjillide, iyjinnine, iylikkixe, iylillige, iymingide, iymirixe, iymissile, iyniddife, iynifice, iynisire, iyrhibbishe, iyrhibirhe, iyshidhdhige, iyshihide, iysihile, iysikire, iysilliye*, iysimmire, iysixixe, iythihire, iythimmite, iythirrife, iythiwwire, iyticibe, iytikile, iytirire, iytsiriye*, iywiddixe, iyzikkire, iyziwwire, izbite, izhire, izikkire, iziwwire, izzikkire.*

For the generic rule of this subcategory, we use as an example the verb *ibbirhe* [to hold, to keep, to arrest, to catch] for which the *SaCoFlexor* module generates the four stems *ibbirh, abbirh, abbarh* and *ibbirh* from which it generates the following flexed forms [note the same stem for both Perfect and Imperfect forms, *ibbirh*].

Table 1: generic rule for C-I verbs starting /e/.

Formula	Flexed form	Rule Id	Conjugation
i-irhe	ibbirhe	PERFEC-1s	PERFECT 1st person singular
ti-irhe	tibbirhe	PERFEC-2s	PERFECT 2nd person singular
yi-irhe	yibbirhe	PERFEC-3sM	PERFECT 3rd person singular masculine
ti-irhe	tibbirhe	PERFEC-3sF	PERFECT 3rd person singular feminine
ni-irhe	nibbirhe	PERFEC-1p	PERFECT 1st person plural
ti-irhin	tibbirhin	PERFEC-2p	PERFECT 2nd person plural
yi-irhin	yibbirhin	PERFEC-3p	PERFECT 3rd person plural

a-irhe	abbirhe	IMPERF-1s	IMPERFECT 1st person singular
ta-irhe	tabbirhe	IMPERF-2s	IMPERFECT 2nd person singular
ya-irhe	yabbirhe	IMPERF-3sM	IMPERFECT 3rd person singular masculine
ta-irhe	tabbirhe	IMPERF-3sF	IMPERFECT 3rd person singular feminine
na-irhe	nabbirhe	IMPERF-1p	IMPERFECT 1st person plural
ta-irhin	tabbirhin	IMPERF-2p	IMPERFECT 2nd person plural
ya-irhin	yabbirhin	IMPERF-3p	IMPERFECT 3rd person plural
a-arho	abbarho	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-arho	tabbarho	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-arho	yabbarho	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-arho	tabbarho	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-arho	nabbarho	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-arhoona	tabbarhoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-arhoona	yabbarhoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
i-irh	ibbirh	IMPERA-2s	IMPERATIVE 2nd person singular
i-irha	ibbirha	IMPERA-2p	IMPERATIVE 2nd person plural

3.1.1 C-I Verbs starting with /i/ and ending with /ye/

The 25 verbs of this subcategory that ends with *"ye"* modify the above-mentioned generic rule in their Subjunctive and Imperative forms. We use as an example the verb *iyfiliye* [to separate, to divide]. The *SaCoFlexor* module generates the four stems *iyfiliy, iyfiliy, ayfalaw*, and *iyfilii*, where the Perfect and Imperfect conjugations follow the generic rule expressed Table 1, while the Subjective substitutes the final */o/* with *'aw'* and the Imperative ends with */ii/* instead of the final phoneme of the generic rule (i.e., /ii/ [iyfil***ii***] instead of /rh/ [ibbi***rh***]). The flexed forms in this case are:

Table 1.1: Additional rules for C-I Verbs starting with /i/ and ending with /ye/

Formula	Flexed form	Rule	Conjugation
a-alaw	ayfalaw	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-alaw	tayfalaw	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-alaw	yayfalaw	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-alaw	tayfalaw	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-alaw	nayfalaw	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-alawoona	tayfalawoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-alawoona	yayfalawoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
i-ii	iyfilii	IMPERA-2s	IMPERATIVE 2nd person singular
i-ii-ya	iyfiliiya	IMPERA-2p	IMPERATIVE 2nd person plural

3.1.2 C-I Verbs starting with /i/ and having less than 3 phonemes

If the verb has only 2 phonemes, the Subjunctive and Imperative conjugations change, and in the case of *ine* [to be (present), to exist, to survive], they become *annaw* and *ennaa* respectively. This subcategory has only the verb *ine* with two phonemes.

Table 1.2: Additional rules for C-I Verbs starting with /i/ having less than 3 phonemes [verb ine]

Formula	Flexed form	Rule	Conjugation
a-naw	annaw	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-naw	tannaw	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-naw	yannaw	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-naw	tannaw	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-naw	nannaw	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-nawoona	nannawoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-nawoona	yannawoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
e-aa	ennaa	IMPERA-2s	IMPERATIVE 2nd person singular
e-aay	ennaay	IMPERA-2p	IMPERATIVE 2nd person plural

3.2 Subcategory 2: C-I verbs starting with the vowel *u.*

In the current Saho Corpus data dictionary, there are no C-I verbs starting with long vowel */uu/*[10], and we have 129 verbs in this subcategory and all of them follow one generic rule with a minor exception of the 12 verbs ending with *"we"*, 3 verbs with the same root (*uluuse* [to knead, to mix], *ulluuse*, and *umluuse*) that has the long vowel /uu/ in their second phoneme, and one verb (*uble* [to see]) with less than three phonemes. These have a slight variation of inflectional system to generate the stems for Subjunctive and Imperative conjugations (see paragraph 3.2.1, 3.2.2, and 3.2.3).

ubbuxute, uble, ucbude, uckure, ucluwe, ucrufe, ucundhuwe, ucussube, udhdhurhe, udluse, udruse, udumdume, uffutute, ufkune, ugshuwe, ujjuure, ukhdume, ukhluqhe, uktube, ukummuse, ukunkune, ukuskuse, ukuxdude, ukuxkuxe, ulluuse, uluuse, umbulluwe, umcube, umcuge, umcure, umfukune, umluuse, umrure, umsuluge, umuccuge, umudhdhuwe, umuggure, umundhuce, umuqquce, umurrufe, umuttuke, umxufuse, umxukume, umxurruse, undhuce, undufule, unkhulluqhe, unkuttube, unkuxkuxe, unqule, unqullufe, unquse, unqutstsuwe, unshurufe, unsuffure, unsulxune, unsummuqhe, untufurhe, untullule, unturjume, uqhlufe, uqhruse, uqqususe, urhuwe, urhxurhe, urkute, urxude, uscrufe, usculuwe, uscundhuwe, uscussube, usfure, ushshucube, ushshucuge, ushshukume, ushshurufe, ushshutuke, ushtuqhule, uskhudume, uskute, uskutube, uskuxdude, usluge, usmundhuce, usmuqhe, usquruse, ussubulluwe, ussucurrufe, ussufukkune, ussukute, ussumurrure, ussuxuttute, usulxune, usxukume, usxurtume, usxuruse, utlule, utsmuqhe, uttukuttube, uttuxuffuse, uttuxurruse, uturjume, uxfuse, uxkume, uxmude, uxmuge, uxruse, uxtume, uxulkuse, uxullufe, uxurtume, uybulluwe, uybuxute, uydumdume, uyduruse, uyfukune, uyfutute, uymurure, uynuqhule, uynuqhuse, uyshuqhule, uyshushshule, uysuluge, uysulxune, uysussule, uytufurhe, uyturjume, uyzuure, uzuure.

Let us take as an example *udruse* [to learn] for the general rule of this subcategory: *SaCoFlexor* generates the four stems *udrus, adrus, adras, udrus,* respectively for Perfect, Imperfect, Subjunctive and Imperative forms, from which it generates the following conjugations.

[10] see note 7

Table 2: Generic rule for C-I verbs starting with the vowel /u/.

Formula	Flexed form	Rule Id	Conjugation
u-use	udruse	PERFEC-1s	PERFECT 1st person singular
tu-use	tudruse	PERFEC-2s	PERFECT 2nd person singular
yu-use	yudruse	PERFEC-3sM	PERFECT 3rd person singular masculine
tu-use	tudruse	PERFEC-3sF	PERFECT 3rd person singular feminine
nu-use	nudruse	PERFEC-1p	PERFECT 1st person plural
tu-usin	tudrusin	PERFEC-2p	PERFECT 2nd person plural
yu-usin	yudrusin	PERFEC-3p	PERFECT 3rd person plural
a-use	adruse	IMPERF-1s	IMPERFECT 1st person singular
ta-use	tadruse	IMPERF-2s	IMPERFECT 2nd person singular
ya-use	yadruse	IMPERF-3sM	IMPERFECT 3rd person singular masculine
ta-use	tadruse	IMPERF-3sF	IMPERFECT 3rd person singular feminine
na-use	nadruse	IMPERF-1p	IMPERFECT 1st person plural
ta-usin	tadrusin	IMPERF-2p	IMPERFECT 2nd person plural
ya-usin	yadrusin	IMPERF-3p	IMPERFECT 3rd person plural
a-aso	adraso	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-aso	tadraso	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-aso	yadraso	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-aso	tadraso	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-aso	nadraso	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-asoona	tadrasoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-asoona	yadrasoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
u-us	udrus	IMPERA-2s	IMPERATIVE 2nd person singular
u-usa	udrusa	IMPERA-2p	IMPERATIVE 2nd person plural

3.2.1 C-I Verbs starting with /u/ and ending with /we/

The 12 verbs of this subcategory that ends with *"we"* modify the above-mentioned generic rule in their Subjunctive and Imperative forms. The verbs are *ucluwe, ucundhuwe, ugshuwe, umbulluwe, umudhdhuwe, unqutstsuwe, urhuwe, usculuwe, uscundhuwe, ussubulluwe, utsmuqhe, uybulluwe.* We use as example the verb *ucluwe* [to escape, to flee, to run away]. The *SaCoFlexor* module generates the four stems *ucluw, acluw, aclaw, and ucluu*, where the Perfect and Imperfect conjugations follow the generic rule [see Table 2], while the Subjunctive substitutes the final /*o*/ with *'aw'* and the Imperative ends with *uu* instead of the final phoneme of the generic rule (i.e., /uu/ [ucl***uu***] for *ucluwe* instead of /r/ [uckur] for *uckure*). The flexed forms in this case are:

Table 2.1: exception rule for C-I verbs starting with the vowel /u/ and ending with "we"

Formula	Flexed form	Rule Id	Conjugation
a-aso	aclaw	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-aso	taclawo	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-aso	yaclawo	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-aso	taclawo	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-aso	naclawo	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-asoona	taclawoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-asoona	yaclawoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
u-uu	ucluu	IMPERA-2s	IMPERATIVE 2nd person singular
u-uuya	ucluuya	IMPERA-2p	IMPERATIVE 2nd person plural

3.2.2 C-I Verbs starting with /u/ with the presence of long vowel /uu/ in the second phoneme

The following three cases (*umluuse* [passive form of *uluuse* = to knead, to mix], *uluuse* [to knead, to mix], and *ulluuse* [middle of *uluuse*]) need specific adjustment where the Imperative form is abridged into short vowel (i.e., *umlus* instead of the predicted *umluus*; *ulus* instead of *uluus*; and *ullus* instead of *ulluus*). The rest of the conjugations follow the generic rule for this subcategory (see Table 2). The example we use here is the verb *uluuse*.

Table 2.2: Exception rule for C-I verbs starting with /u/ and have long vowel /uu/ in the second phoneme

Formula	Flexed form	Rule Id	Conjugation
u-us	ulus	IMPERA-2s	IMPERATIVE 2nd person singular
u-usa	ulusa	IMPERA-2p	IMPERATIVE 2nd person plural

Further data collected from Saho speaking still confirms the correct version should be following the generic rule, meaning Imperative form of these verbs should be *umluus, uluus,* and *ulluus*[11].

3.2.3 C-I Verbs starting with /u/ and having less than 3 phonemes

Exception: If the verb has only 2 phonemes, the Subjunctive and Imperative conjugations change, and in the in case of *uble* [to see], they become respectively *able* and *ubul.* This category has only the verb *uble* with two phonemes.

Table 4.3: Exception C-I Verbs starting with /u/ and having less than 3 phonemes [Uble]

Formula	Flexed form	Rule	Conjugation
u-e	uble	PERFEC-1s	PERFECT 1st person singular
tu-e	tuble	PERFEC-2s	PERFECT 2nd person singular
yu-e	yuble	PERFEC-3sM	PERFECT 3rd person singular masculine
tu-e	tuble	PERFEC-3sF	PERFECT 3rd person singular feminine
nu-e	nuble	PERFEC-1p	PERFECT 1st person plural
tu-in	tublin	PERFEC-2p	PERFECT 2nd person plural
yu-in	yublin	PERFEC-3p	PERFECT 3rd person plural
a-e	able	IMPERF-1s	IMPERFECT 1st person singular
ta-e	table	IMPERF-2s	IMPERFECT 2nd person singular
ya-e	yable	IMPERF-3sM	IMPERFECT 3rd person singular masculine
ta-e	table	IMPERF-3sF	IMPERFECT 3rd person singular feminine
na-e	nable	IMPERF-1p	IMPERFECT 1st person plural
ta-in	tablin	IMPERF-2p	IMPERFECT 2nd person plural
ya-in	yablin	IMPERF-3p	IMPERFECT 3rd person plural
a-o	ablo	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-o	tablo	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-o	yablo	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-o	tablo	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-o	nablo	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-oona	tabloona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-oona	yabloona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
u-ul	ubul	IMPERA-2s	IMPERATIVE 2nd person singular
u-ula	ubula	IMPERA-2p	IMPERATIVE 2nd person plural

[11] Conversation with Saho Speaking community members in London of 30th October 2022.

3.3 Subcategory 3: C-I verbs starting with vowel *e.*
We differentiate the case for the long vowel */ee/* from the short vowel */e/* (i.e., verbs starting with /*e/* followed by a consonant). For the first case, let us take as example the verb *eebexe* [to sell]. The *SaCoFlexor* module generates in this case the four stems *eebex, aabix, aabax, ibix* to generate the following flexed forms.

Table 3. Generic rule for C-I verbs starting with vowel /e/.

Formula	Flexed form	Rule	Conjugation
ee-exe	eebexe	PERFEC-1s	PERFECT 1st person singular
tee-exe	teebexe	PERFEC-2s	PERFECT 2nd person singular
yee-exe	yeebexe	PERFEC-3sM	PERFECT 3rd person singular masculine
tee-exe	teebexe	PERFEC-3sF	PERFECT 3rd person singular feminine
nee-exe	neebexe	PERFEC-1p	PERFECT 1st person plural
tee-exin	teebexin	PERFEC-2p	PERFECT 2nd person plural
yee-exin	yeebexin	PERFEC-3p	PERFECT 3rd person plural
aa-ix	aabix	IMPERF-1s	IMPERFECT 1st person singular
taa-ixe	taabixe	IMPERF-2s	IMPERFECT 2nd person singular
yaa-ixe	yaabixe	IMPERF-3sM	IMPERFECT 3rd person singular masculine
taa-ixe	taabixe	IMPERF-3sF	IMPERFECT 3rd person singular feminine
naa-ixe	naabixe	IMPERF-1p	IMPERFECT 1st person plural
taa-ixin	taabixin	IMPERF-2p	IMPERFECT 2nd person plural
yaa-ixin	yaabixin	IMPERF-3p	IMPERFECT 3rd person plural
aa-axo	aabaxo	SUBJUN-1s	SUBJUNCTIVE 1st person singular
taa-axo	taabaxo	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
yaa-axo	yaabaxo	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
taa-axo	taabaxo	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
naa-axo	naabaxo	SUBJUN-1p	SUBJUNCTIVE 1st person plural
taa-axoona	taabaxoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
yaa-axoona	yaabaxoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
i-ix	ibix	IMPERA-2s	IMPERATIVE 2nd person singular
i-ixa	ibixa	IMPERA-2p	IMPERATIVE 2nd person plural

This category has a total of 12 verbs in the corpus. The following 10 verbs follow the general rules of this subcategory. The only two verbs identified with this category that do not follow the same rule are those with two phonemes (*eelle* and *eerre:* see for these paragraph 3.3.1*)*.

eebexe, eecete, eedece, eefece, eeleme*, eelle***, eemene**, eemere*, eerhege, eerre***, eexete, eezeze**

(*) if the first phoneme is *d, l, m* and *z*, then the imperative 2nd person follows the rule of *eedece* (i.e. *eedec*), otherwise it follows the rule of *eebex* (i.e. *ibix*).
(**) the imperative 2nd person singular follows both the rules and can be both *imin* and *eemen*.
(***) verbs with two phonemes [see this exception in paragraph 3.3.1].

3.3.1 C-I verbs starting with /e/ and with less than 3 phonemes
The two above mentioned verbs (*eelle* and *eerre*) need to be managed separately. The following is the rule for conjugation (example *eelle*):

Table 3.1 Rule for C-I verbs starting with vowel /e/ and with only two phonemes.

Formula	Flexed form	Rule	Conjugation
ee-le	eelle	PERFEC-1s	PERFECT 1st person singular
tee-e	teelle	PERFEC-2s	PERFECT 2nd person singular
yee-e	yeelle	PERFEC-3sM	PERFECT 3rd person singular masculine
tee-e	teelle	PERFEC-3sF	PERFECT 3rd person singular feminine
nee-e	neelle	PERFEC-1p	PERFECT 1st person plural
tee-in	teellin	PERFEC-2p	PERFECT 2nd person plural
yee-in	yeellin	PERFEC-3p	PERFECT 3rd person plural
aa-e	aalle	IMPERF-1s	IMPERFECT 1st person singular
taa-e	taalle	IMPERF-2s	IMPERFECT 2nd person singular
yaa-e	yaalle	IMPERF-3sM	IMPERFECT 3rd person singular masculine
taa-e	taalle	IMPERF-3sF	IMPERFECT 3rd person singular feminine
naa-e	naalle	IMPERF-1p	IMPERFECT 1st person plural
taa-in	taallin	IMPERF-2p	IMPERFECT 2nd person plural
yaa-in	yaallin	IMPERF-3p	IMPERFECT 3rd person plural
aa-o	aallo	SUBJUN-1s	SUBJUNCTIVE 1st person singular
taa-o	taallo	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
yaa-o	yaallo	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
taa-o	taallo	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
naa-o	naallo	SUBJUN-1p	SUBJUNCTIVE 1st person plural
taa-oona	taalloona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
yaa-oona	yaalloona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
ee-aa	eellaa	IMPERA-2s	IMPERATIVE 2nd person singular
ee-aanta	eellaanta	IMPERA-2p	IMPERATIVE 2nd person plural

3.3.2 C-I verbs starting with /e/ but not /ee/.

In the case of a verb starting with the vowel /*e/* followed by a consonant, we have 106 C-I verbs. Subdivided into 3 subclassifications, the flexor application generic rule successfully generates the flexing forms for the following 106 verbs with only 13 cases that need to be treated separately: 7 of these are verbs with less than 3 phonemes: *edee, ekke, elee, emce, erde, erhxe, eshshe, ewce, eylee.*

eddebbe, edebbe, edee, edhdheerhe, edhdheexe, egeeceze, egeelebe, egeerece, eggerre, ekette, ekeye, ekke*, ekkele, elee*, eleeze, elleeze, emce*, emcelle, emeete, emege, emerre, emfeere, emlelekke, emmeceeceze, emmeege, emmeezene, emmenge*, emmeseesele, emneeberre, emneesexe, emreredde, emrocoocoze, emseene, emsemeemece, emsheeqhele, emweesebe, emxeesebe, emzeegexe, endedebbe, endedeedebbe, endeewe, endhewwe, enebe, eneye, engeece, engeelebe, engele, enkeere, enkeexede, enqeese, enzeewere, eqheese, eqheeweme, erde*, ereeqhebe, erheerhe, erhxe, eseemexe, esgeeceze, esgeegelle, esgeelebe, esgelle, eshshe*, eskeere, eskette, esseece, essekette, essexeedere, esxeedere, esxeesebe, esxesse, eththeecese, eththeehere, etteelee*, ettemege, ewce*, exeesebe, eybeereke, eydedebbe, eydedeedebbe, eydeewe, eyfeere, eylee*, eylelee*, eylelekke, eymeceeceze, eymeece, eymeezene, eymenge*, eymeseesele, eymexeedere, eynebe, eyreredde, eyrheerhe, eyrhewwe, eyseelexe, eyseene, eysemeemece, eysheeqhele, eyteelee*, eytheehere, eyweesebe, eyweesele, eyyeece, eyzeegexe, ezeekere.*

3.3.2.1 Subclassification 1: the generic rule for verbs starting with */e/* and not */ee/*.
Total verbs 96. We use the verb *egeeceze* [to transfer oneself, to move]. The following morphological conjugations are produced:

Table 3.2 Generic rule for C-I verbs starting with vowel /e/ and with only two phonemes.

Formula	Flexed form	Rule	Conjugation
e-le	egeeceze	PERFEC-1s	PERFECT 1st person singular
te-e	tegeeceze	PERFEC-2s	PERFECT 2nd person singular
ye-e	yegeeceze	PERFEC-3sM	PERFECT 3rd person singular masculine
te-e	tegeeceze	PERFEC-3sF	PERFECT 3rd person singular feminine
ne-e	negeeceze	PERFEC-1p	PERFECT 1st person plural
te-in	tegeecezin	PERFEC-2p	PERFECT 2nd person plural
ye-in	yegeecezin	PERFEC-3p	PERFECT 3rd person plural
a-e	ageeceze	IMPERF-1s	IMPERFECT 1st person singular
ta-e	tageeceze	IMPERF-2s	IMPERFECT 2nd person singular
ya-e	yageeceze	IMPERF-3sM	IMPERFECT 3rd person singular masculine
ta-e	tageeceze	IMPERF-3sF	IMPERFECT 3rd person singular feminine
na-e	nageeceze	IMPERF-1p	IMPERFECT 1st person plural
ta-in	tageecezin	IMPERF-2p	IMPERFECT 2nd person plural
ya-in	yageecezin	IMPERF-3p	IMPERFECT 3rd person plural
a-o	agaacazo	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-o	tagaacazo	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-o	yagaacazo	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-o	tagaacazo	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-o	nagaacazo	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-oona	tagaacazoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-oona	yagaacazoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
e-	egeecez	IMPERA-2s	IMPERATIVE 2nd person singular
e-a	egeeceza	IMPERA-2p	IMPERATIVE 2nd person plural

3.3.2.2 Subclassification 2: rule for verbs starting with */e/* and not */ee/* and have < 2 phonemes.
Total verbs 7. We use the verb *ewce* [to go out]. The following morphological conjugations are produced:

Table 3.3 Rule for C-I verbs starting with vowel /e/ and with only two phonemes.

Formula	Flexed form	Rule	Conjugation
e-le	ewce	PERFEC-1s	PERFECT 1st person singular
te-e	tewce	PERFEC-2s	PERFECT 2nd person singular
ye-e	yewce	PERFEC-3sM	PERFECT 3rd person singular masculine
te-e	tewce	PERFEC-3sF	PERFECT 3rd person singular feminine
ne-e	newce	PERFEC-1p	PERFECT 1st person plural
te-in	tewcin	PERFEC-2p	PERFECT 2nd person plural
ye-in	yewcin	PERFEC-3p	PERFECT 3rd person plural
a-e	awce	IMPERF-1s	IMPERFECT 1st person singular
ta-e	tawce	IMPERF-2s	IMPERFECT 2nd person singular
ya-e	yawce	IMPERF-3sM	IMPERFECT 3rd person singular masculine
ta-e	tawce	IMPERF-3sF	IMPERFECT 3rd person singular feminine
na-e	nawce	IMPERF-1p	IMPERFECT 1st person plural
ta-in	tawcin	IMPERF-2p	IMPERFECT 2nd person plural
ya-in	yawcin	IMPERF-3p	IMPERFECT 3rd person plural

a-o	awco	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-o	tawco	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-o	yawco	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-o	tawco	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-o	nawco	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-oona	tawcoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-oona	yawcoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
e-	ewec	IMPERA-2s	IMPERATIVE 2nd person singular
e-a	eweca	IMPERA-2p	IMPERATIVE 2nd person plural

3.3.2.3 Subclassification 3: The following 6 verbs of this subcategory need to develop a separate rule in this category, and they are: *emmenge, emrocoocoze, eteellee, eylelee, eymenge, eyteelee.*

3.4 Subcategory 4: C-I verbs starting with the vowel *o.*

In this subcategory, we differentiate the case for of long vowel *'oo'* than single *'o'*. For the first case, we use the example *ooboke* (to be born). The *SaCoFlexor* module calculates the four stems *oobok, aabuk, aabak, ubuk* that generate conjugations of following forms.

Table 4: C-I verbs starting with /o/.

Formula	Flexed form	Rule	Conjugation
oo-oke	ooboke	PERFEC-1s	PERFECT 1st person singular
too-oke	tooboke	PERFEC-2s	PERFECT 2nd person singular
yoo-oke	yooboke	PERFEC-3sM	PERFECT 3rd person singular masculine
too-oke	tooboke	PERFEC-3sF	PERFECT 3rd person singular feminine
noo-oke	nooboke	PERFEC-1p	PERFECT 1st person plural
too-okin	toobokin	PERFEC-2p	PERFECT 2nd person plural
yoo-okin	yoobokin	PERFEC-3p	PERFECT 3rd person plural
aa-uke	aabuke	IMPERF-1s	IMPERFECT 1st person singular
taa-uke	taabuke	IMPERF-2s	IMPERFECT 2nd person singular
yaa-uke	yaabuke	IMPERF-3sM	IMPERFECT 3rd person singular masculine
taa-uke	taabuke	IMPERF-3sF	IMPERFECT 3rd person singular feminine
naa-uke	naabuke	IMPERF-1p	IMPERFECT 1st person plural
taa-ukin	taabukin	IMPERF-2p	IMPERFECT 2nd person plural
yaa-ukin	yaabukin	IMPERF-3p	IMPERFECT 3rd person plural
aa-ako	aabako	SUBJUN-1s	SUBJUNCTIVE 1st person singular
taa-ako	taabako	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
yaa-ako	yaabako	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
taa-ako	taabako	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
naa-ako	naabako	SUBJUN-1p	SUBJUNCTIVE 1st person plural
taa-akoona	taabakoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
yaa-akoona	yaabakoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
u-uk	ubuk	IMPERA-2s	IMPERATIVE 2nd person singular
u-uka	ubuka	IMPERA-2p	IMPERATIVE 2nd person plural

This subcategory totals 13 verbs in the corpus. The following 11 verbs follow the general rules of this subcategory. The only two verbs identified with this subcategory but that do not follow the same rule are *oobbe* and *oome*.

ooboke, oocobe, oocoge, oocore, oofowe, oogore, ookome, ooqhoce, oorofe, oosole, ootoke.*

(*) The verb *oofowe* can be pronounced as *oofoye*, and therefore Subjunctive becomes *aafaw* [because of the exceptional rules for verbs ending with 'ye', see Table 1.1], and Imperative could be both *oofoo* and *ufuu* [see Table 1.1].

3.4.1 C-I verbs starting with /o/ but not /oo/.
In the case of the verb starts with */o/* followed by a constant, and taking as example the verb *omroocod* (to be agitated, to wriggle), the following stems are calculated by the *SaCoFlexor (omroocod, amroocod, amraacad, omroocod)* from which the below list of flexed forms are generated:

Table 4.1: Rule for C-I verbs starting with /o/ but not /oo/.

Formula	Flexed form	Rule	Conjugation
o-ode	omroocode	PERFEC-1s	PERFECT 1st person singular
to-ode	tomroocode	PERFEC-2s	PERFECT 2nd person singular
yo-ode	yomroocode	PERFEC-3sM	PERFECT 3rd person singular masculine
to-ode	tomroocode	PERFEC-3sF	PERFECT 3rd person singular feminine
no-ode	nomroocode	PERFEC-1p	PERFECT 1st person plural
to-odin	tomroocodin	PERFEC-2p	PERFECT 2nd person plural
yo-odin	yomroocodin	PERFEC-3p	PERFECT 3rd person plural
a-ode	amroocode	IMPERF-1s	IMPERFECT 1st person singular
ta-ode	tamroocode	IMPERF-2s	IMPERFECT 2nd person singular
ya-ode	yamroocode	IMPERF-3sM	IMPERFECT 3rd person singular masculine
ta-ode	tamroocode	IMPERF-3sF	IMPERFECT 3rd person singular feminine
na-ode	namroocode	IMPERF-1p	IMPERFECT 1st person plural
ta-odin	tamroocodin	IMPERF-2p	IMPERFECT 2nd person plural
ya-odin	yamroocodin	IMPERF-3p	IMPERFECT 3rd person plural
a-ado	amraacado	SUBJUN-1s	SUBJUNCTIVE 1st person singular
ta-ado	tamraacado	SUBJUN-2s	SUBJUNCTIVE 2nd person singular
ya-ado	yamraacado	SUBJUN-3sM	SUBJUNCTIVE 3rd person singular masculine
ta-ado	tamraacado	SUBJUN-3sF	SUBJUNCTIVE 3rd person singular feminine
na-ado	namraacado	SUBJUN-1p	SUBJUNCTIVE 1st person plural
ta-adoona	tamraacadoona	SUBJUN-2p	SUBJUNCTIVE 2nd person plural
ya-adoona	yamraacadoona	SUBJUN-3p	SUBJUNCTIVE 3rd person plural
o-od	omroocod	IMPERA-2s	IMPERATIVE 2nd person singular
o-oda	omroocoda	IMPERA-2p	IMPERATIVE 2nd person plural

We have 25 verbs in this category. The following 21 verbs follow the above generic rules. The only four verbs that are irregular in this case are *omobbe, oshshoobbe, ottoxoye, oxoye*. Furthermore, among the 25 listed, the verbs with * below needs to be checked in one or two flexed forms.

obboohose, ogoofofe, oloole, omgoofofe, ommokookome, ommoofowe*, ommookome, omroocode, oqqoomote, oshshoome*, oshshoome, osoome, osqoome, otoobe*, ottoocore**, ottoxoye, oxoye, oyloole, oymookome, oyqhoce**, oyroocode*

(**) The verb *oyqhoce* has two forms, *oyqhoce* and *uyqhoce*.

4. Conclusion

In this paper, we have developed an NLP method and written an application using it as part of the Saho Corpus tools, to automatically generate the inflectional forms of the C-I verbs (using the terminology of Banti and Vergari, 2005) present in the current Corpus database. A total of 586 verbs have been classified, and for each of them the morphological conjugations of Perfect, Imperfect and Subjunctive forms have been introduced, for the 1st, 2nd, and 3rd person in the singular and in the plural, and furthermore for the third person, have been introduced both masculine and feminine forms. Only 7 verbs (out of the 586) needed to manually produce the conjugation forms, therefore the module managed automatically 98.80% of the C-I verbs present in the current database. A total of 13,455 words (12870 new flexed forms plus 586 already present) have been tagged correctly and codified for further use of the Saho Corpus tools on search, lemmatization, collocation, and concordance utilities. The application's codified output data fed into the Saho Corpus database, to enhance NLP functions, including word identification, concordance, and collocation listing, as well as word usage frequency functions. In the following paragraph 4.1 a list of typos in verb conjugations corrected by the module is listed.

4.1 Spelling error or need for normalization on the dictionary

The major reference to check and verify the correctness of the data collected and/or calculated has been the basic dictionary of the Saho Language by [Vergari & Vegari, 2007]. The following type errors has been noted in the dictionary for the conjugation of the verbs analysed in this paper. They are reported here to be checked for the next edition by the curators. For each verb it is indicated which conjugation form that has some issues to check (Perfect, Imperfect, Subjunctive and Imperative), and if necessary, the case (1st, 2nd, 3rd, plural, and genre). The page number is not reported as the reference has been done a PDF edition shared directly by the curators.

Verb	Conjugation	Correct value	Dictionary
ottoocore	IMPE	attoocore	att***aa***c***u***re
unqullufe	SUBJ	anqallafo	anqallaf***a***
uqhlufe	SUBJ	aqhlafo	aqhlaf***a***
uyshuqhule	IMPV	uyshu***qh***ul	uyshu***hq***ul
umuccuge	SUBJ	am***a***ccago	am***u***ccago
icrhiye	SUBJ	acrhaw	acrhaw***o***
imjillide	IMPV	imjillid	ijillid [missing m]
iyfirixe	IMPV	iyfirix	iyfrix [missing i]
imwirrishe	IMPV	imwirrish	imwirish [missing r]
isgiribe	IMPV	isgirib	isgir***u***b
igribe	IMPV	igrib	igr***u***b
imwirrishe	IMPE	amwirrishe	***i***mwirrishe
iskire	SUBJ	askaro	ask***i***ro
imibbixe	IMPE	amibbixe	am***a***bbixe
ishshimine	IMPE	ashshimine	ashsh***a***mine
iskimile	IMPE	askimile	askimil***a***

Bibliography

AXMADSACAD MAXAMMAD CUMAR, Giorgio BANTI, and Moreno VERGARI (2014), "Saho Islamic poetry and other literary genres in Ajami script", Oral presentation at *"Manuscripts & c. in the Horn of Africa"*, Asien-Afrika-Institut, Hamburg, 17-19 July 2014.

BANTI, Giorgio, and Moreno VERGARI (2005), "A sketch of Saho grammar", *Journal of Eritrean Studies*, IV: 100-131.

BANTI, Giorgio, and Moreno VERGARI (2023), "A sketch of Saho grammar", *Oxford Handbook of Ethiopian Languages*, forthcoming.

CHEFENA HAILEMARIAM (2002), *Language and Education in Eritrea*, Amsterdam: Aksant.

D'ABBADIE, Antoine (1843), "Sur la langue saho". *Journal Asiatique* 4/2, 108-18.

ESAYAS TAJEBE (2005), *Descriptive Grammar of Saaho*, Unpublished PhD thesis. Addis Ababa University.

GHIRMAI NEGASH (1999), *A History of Tigrinya Literature in Eritrea. The oral and the written 1890-1991*, Leiden: CNWS.

HAYWARD, Richard J. (1979), "Some inferences from irregular Imperative form in Saho". *Israel Oriental Studies* 9:245-257.

JAMA MUSSE JAMA (2016), *A Syntactically Annotated Corpus of Somali Literature*, Oriental University of Naples, PhD thesis [unpublished]. See www.somalicorpus.com.

JAMA MUSSE JAMA (2023), *Saho Corpus: state of the art, and tools for linguistic analysis* (to appear on Dhaxalreeb Vol 20, No. 1).

MINISTRY OF EDUCATION (2011), *Saahot Luqha, 6ya fasleh, Darasat kitaab*, Department of General Education, Curriculum Planning and Development Division, ASMARA.

MINISTRY OF EDUCATION (2015), *Saahot xarfi kitaab afti fasleh*, Department of General Education, Curriculum Planning and Development Division, Sabur Printing Services, Asmara.

MINISTRY OF EDUCATION (2016), *Saahot Waani, 3it fasleh, Darasat kitaab*, Department of General Education, Curriculum Planning and Development Division, Asmara.

REDIE BEREKETEAB (2010), "The Politics of Language in Eritrea: Equality of Languages Vs. Bilingual Official Language Policy", *African and Asian Studies*, 9 (1-2), 149-190. https://doi.org/10.1163/156921010X491308.

REINISCH, Leo (1877), "Studien über Ost-Afrika - Das Saho-Volk", *Österreichische Monatsschrift für den Orient*, III: 65-73.

REINISCH, Leo (1878a), "Die Saho Sprache", *Zeitschr. der Deutschen Morgenländischen Gesellschaft*, XXXII: 415 - 464.

REINISCH, Leo (1878b), "Die Sprache der Irob-Saho in Abessinien", *Sitzungsb. der Phil.-Hist. Classe der Kaiserl.* Akademie der Wissenschaften, XC: 89-142.

REINISCH, Leo (1889), *Die Saho-Sprache*, Vol. 1 Texte, Wien: Alfred Hölder.

REINISCH, Leo (1890), *Die Saho-Sprache*, Vol. 2 Wörterbuch, Wien: Alfred Hölder.

SALT, Henry (1814), *A Voyage to Abyssinia and travels into the interior of that country, executed under the orders of the British government in the years 1809 & 1810*, London: F.C. And J. Rivington By W. Bulmer and Co.

TEKLE ABRAHA (1995), "Alif, Ba, Ta and A, B, C: Writing Eritrea", *Eritrea Profile*, February 25, 1995.

TIRSIT YETBAREK and JAMA MUSSE JAMA (2022), "Extension of the Somali Corpus structure to the Omo-Tana branch of the Cushitic languages family", Oral presentation at *"International Workshop on the Archaeology, History and Related Issues of the Somali-speaking countries"*, Oriental University of Naples, 15 September 2022.

VERGARI, Moreno and Roberta VERGARI (2003), *A basic Saho-English-Italian dictionary*, Asmara: Sabur Printing Services. [Revised edition 2007].

VOIGT, Rainer Maria (1975), "Bibliographie des Saho-cAfar". *Africana Marburgensia* 8/1, 53-63.

WELMERS, William Evert (1952), "Notes on the structure of Saho". *Word* 8, 145-62, 236-51.

ZABORSKI, Andrzej (1987), "ᶜAfar-Saho linguistics. An overview", in H.G. Mukarovsky (ed.), *Leo Reinisch. Werk und Erbe*, 85-95. Vienna.