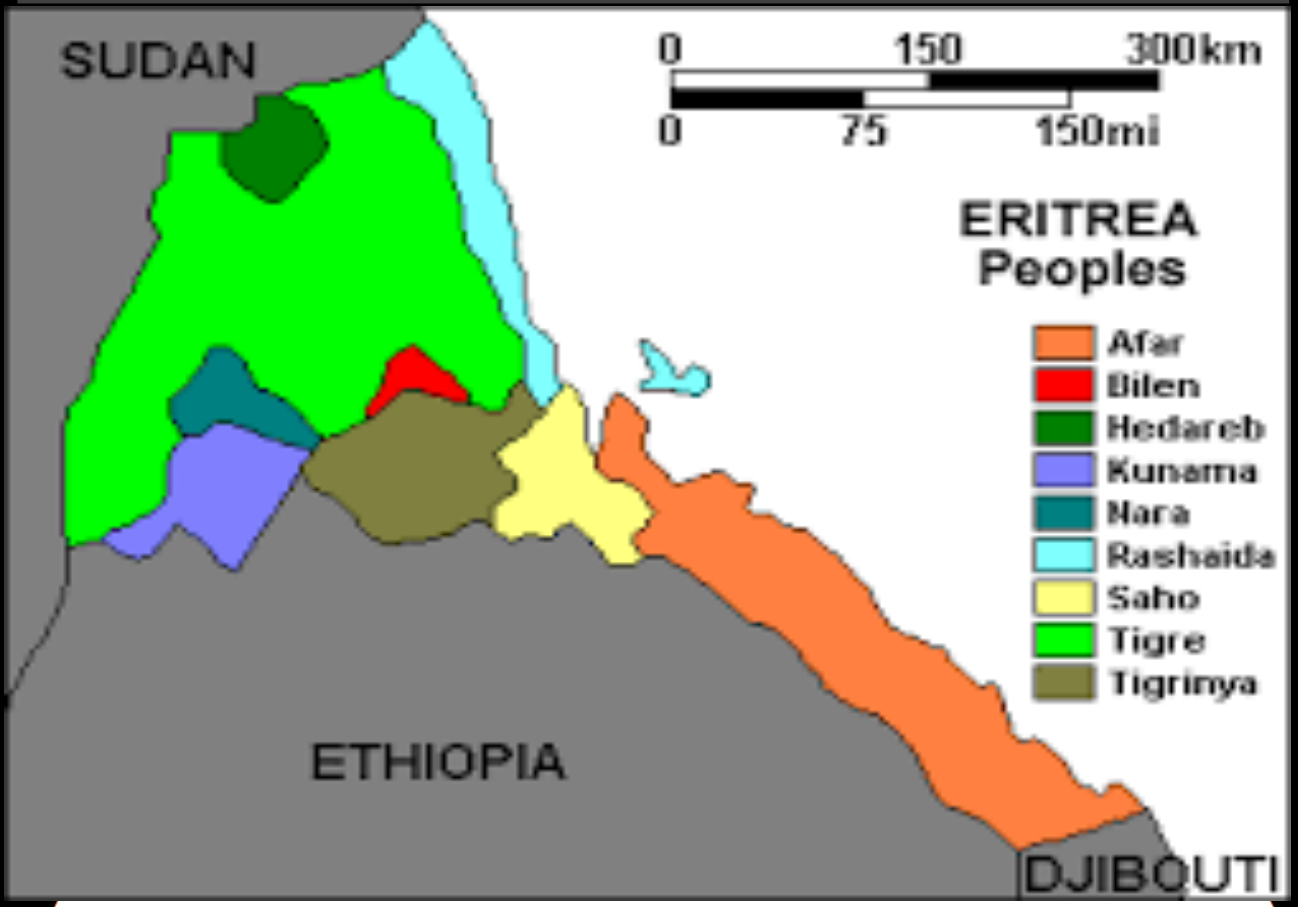
SAMTRAMOD
SAAHO MACHINE TRANSLATION MODEL - 1.0

Dr Jama Musse Jama, PhD
Hargeysa Cultural Centre
Somaliland Centre for African Studies
Redsea Cultural Foundation

4th Hybrid Cushitic Conference - Leiden 2025, December 11-12
Linguistic History of East Africa - Leiden University
12th Decembre 2025

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

SAAHO PEOPLE



Afar, Beja (Hedareb), Bilen, Kunama, Nara, Rashaida, Saaho, Tigre, Tigrinya

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO PEOPLE / LANGUAGE

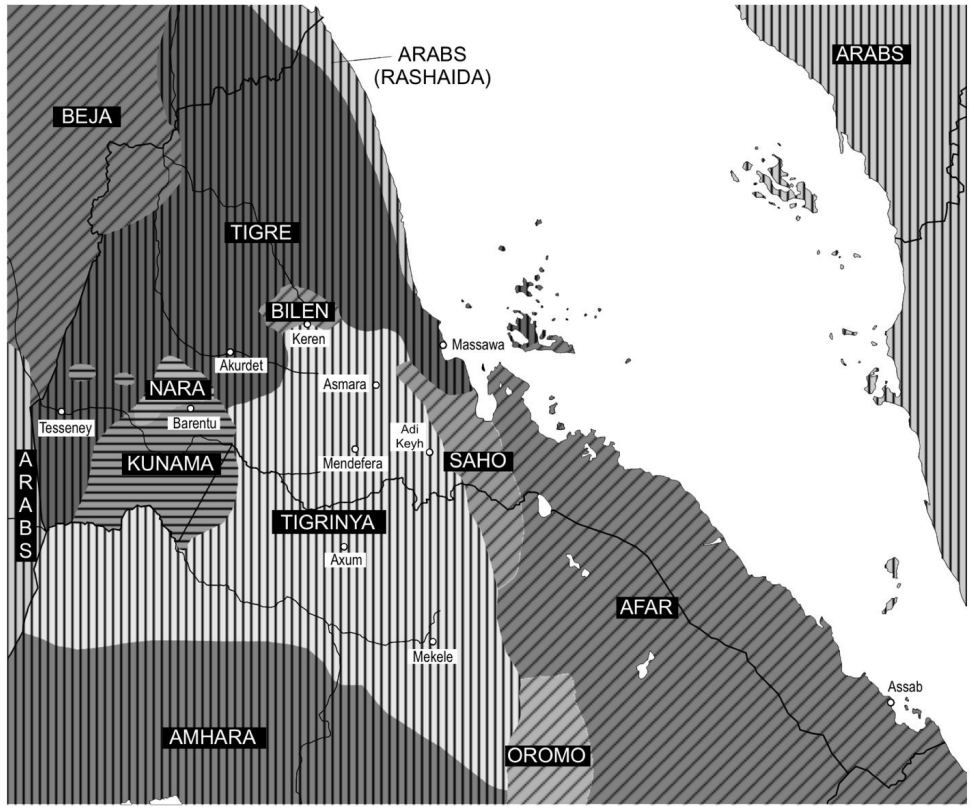


FIGURE 2 - Approximative distribution of ethnic groups and languages in Eritrea and bordering regions (adapted and modified from: European Asylum Support Office (2015) *Eritrea Country Focus*, p. 13. Luxembourg: Publications Office of the European Union).

Semitic languages | Cushitic languages | Nilo-Saharan languages

Banti & Vergari, 2017

- THE LANGUAGE
- SAAHO CORPUS
- MT FOR CUSHITIC FAMILY
- SAAHO DATA SOURCES
- USER INTERFACE
- SAMTRAMOD - EVALUATION
- CONCLUSIONS

SAAHO LANGUAGE

Saaho is a Lowland East Cushitic language that is spoken in central and southeastern Eritrea by the Saaho community, as well as in the neighbouring region of Tigray, Ethiopia by the Irob community.

It can be divided into three major dialects - Northern Saaho, Central Saaho, and Southern Saaho

Saahot waani, Saahot lux'a, Saahot k'onk'a, Shaho, Shoho, and Saho

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO LANGUAGE

It is estimated that Saaho has about 250,000 speakers worldwide (Banti and Vergari 2024).

The Saaho language has been transcribed in early 1800s, and soon after described, and studied primarily by non-Saaho native linguists and travellers.

There is Diasporic factor

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO LANGUAGE

Then over the last four decades Zaborski (1987); Banti and Vergari (2010, 2023, 2024); Hayward (1979; 1983; 1997), Hayward & Orwin (1991); Lamberti (1990) and Morin (1995; 1999; 2006; 2015) documented and described, also engaging with Saaho native speakers.

Recent attention from the scholarship

THE LANGUAGE

- SAAHO CORPUS
- MT FOR CUSHITIC FAMILY
- SAAHO DATA SOURCES
- USER INTERFACE
- SAMTRAMOD - EVALUATION
- CONCLUSIONS

SAAHO LANGUAGE

More recently, there has been an increase in the development of Natural Language Processing (NLP) modules for Saaho, including analyses of word frequencies, collocations, concordances, and annotated corpus data.

Jama Musse (2021, 2022, 2024, 2025); Di Manno et al., (2024)

Somali Corpus and Universal Dependencies Dataset

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

SAAHO CORPUS

www.sahocorpus.com

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO CORPUS

The Saho Corpus is a comprehensive repository developed by the present author in 2020.

It now contains over one million tagged words in grammatically validated texts, and provides tools for searching and analysis.

Orthography: Used only Latin (1983, 1996). Missing literature previously written using an orthography based on the Fidel script in Ethiopia.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

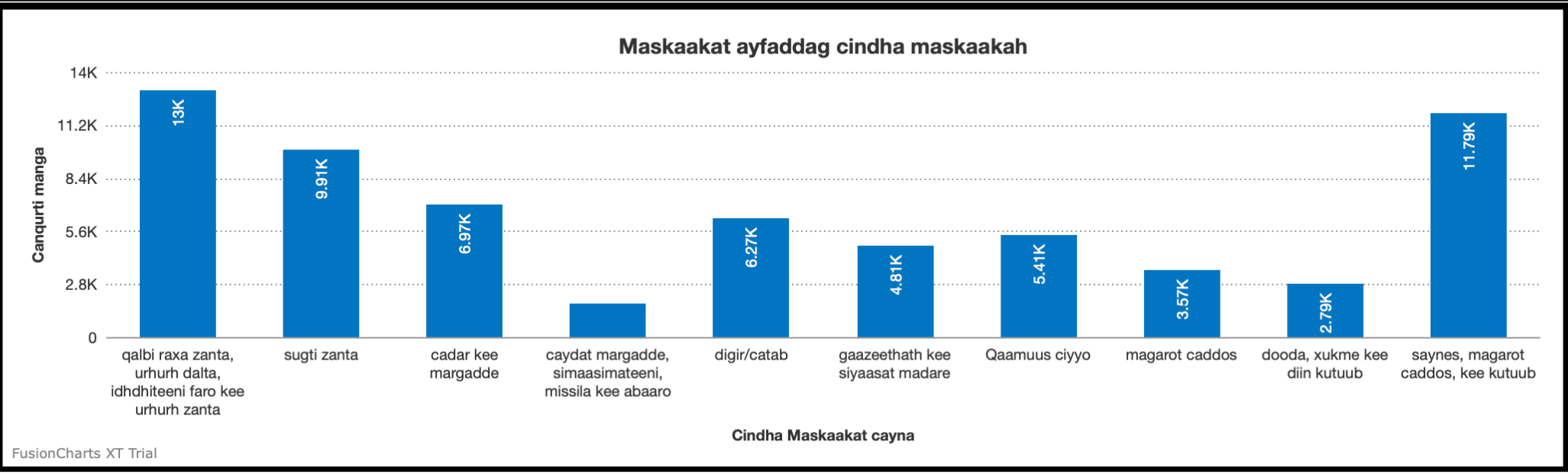
USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

Saahot Afhi Maskaaka

Canqurko 1 miliyoonko bukalle takkem yemweesebe | 440 takkem dokkumentitko barhsimeh tuumaarime | amot canqurko 4700 bukalle takkem geytime yimfillice canqurko 20 shixtiya takkem yewce | 4500ko bukalle yakke canqur 2 luqhah yunturjume



orob: geli-canqara: Maskaakadde orob Maanafo ab geli canqara/yatneefece tiyahi muyrhurhug biyashshiite

Saahot Maskaaka cakanat darajalle kinni tanem. Tayek basoh, yatneefece tiyahi muyrhurhug kee geli-canqarahi araxah bisa kinni adde orobto dhictam. Niiroh laakin, umman tiyah fakkot takko kinni. Ta cakanat hoorudde tangalo fadhdhemko, ta gana tane xogoda tetneefeceh 'Maanafo gin', hido lel ta kataa araxah farto nolle uktub: info@sahocorpus.com cc: sahocorpus@gmail.com

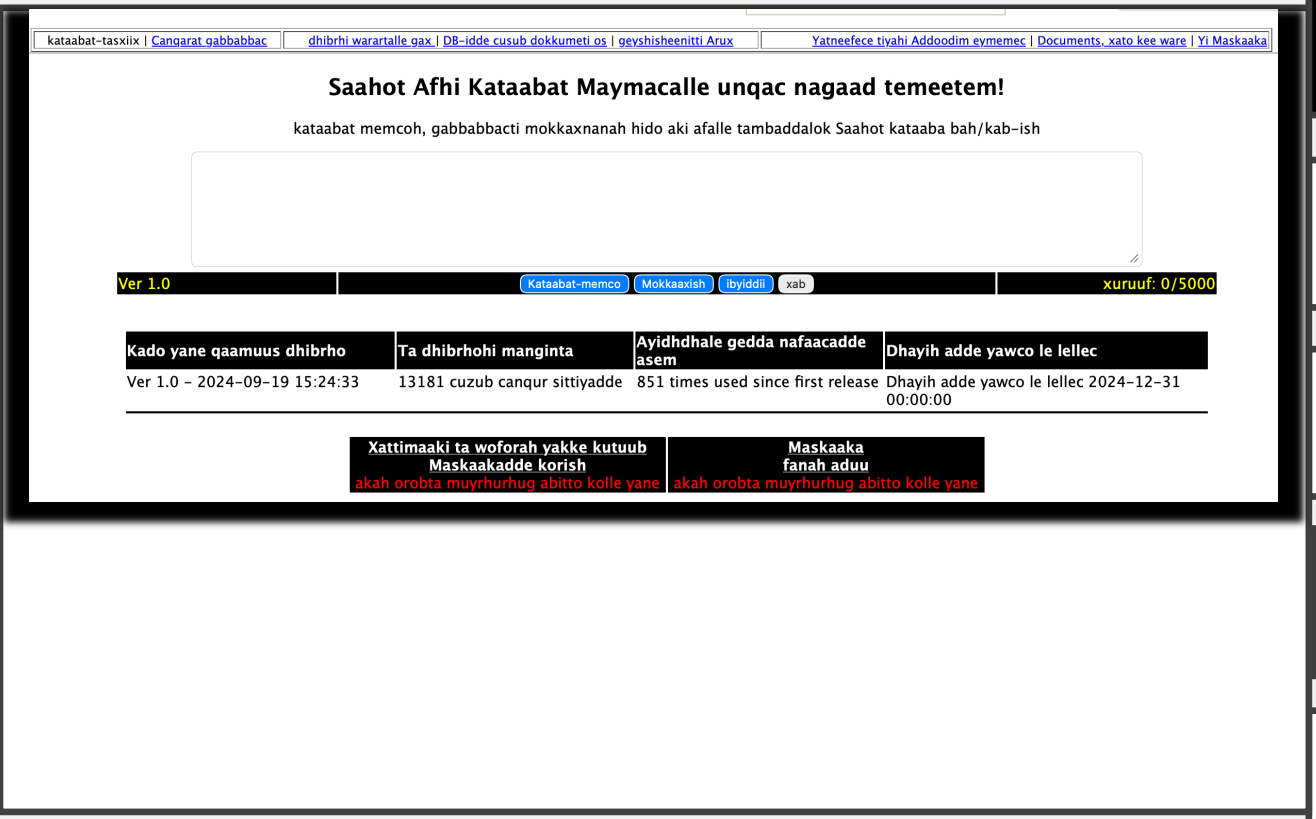
REDSEA CULTURAL FOUNDATION www.redsea-online.org

A project designed and directed by
Dr. Jama Musse Jama, PhD
Redsea Cultural Foundation
website | email: jama@redsea-online.org

With the collaboration of
Prof Giorgio Banti
Oriental Univesity of Naples
website | email:gibanti@unior.it

Special thanks to Moreno Vergari, Ethnoréma, Ahmedsaad Mohammed Omer and Saho Speaking Community in London

SAAHO CORPUS



www.sahocorpus.com

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO CORPUS

The corpus has a user interface that offers concordance, collocations, and word frequency information for predefined subcorpora of literature.

A community-driven project, with linguists, lexicographers, and cultural experts from the Saaho community, as well as non-Saaho scholars working on the language

www.sahocorpus.com

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO CORPUS

The Saho Corpus is part of a broader set of corpora for Cushitic languages, including Somali (2016), the Omo-Tana family of languages (Rendille, Bayso, Dhaasanac Maay, Arbore, and Boni in addition to Somali) (2024-2025), and Oromo (under construction expected completion in April 2026).

www.somalicorpus.com

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO CORPUS

These corpora are all built on similar technological platforms based on the *Kaydka Af Soomaaliga* (KAF), Somali Corpus.

Not only has the structure and interface typical of standard corpus building and core functionalities (such as concordance and collocations), but it also includes a primary monolingual dictionary.

www.somalicorpus.com

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO CORPUS

This dictionary serves as a starting point for search and data retrieval, providing users with the root words, their definitions, grammatical, and morphological notations, and translations into other languages.

For example, easily POS Tagger availability.

www.somalicorpus.com

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

NLP – CUSHITIC FAMILY

Languages are defined as Low Resourced (or Scarce) when there is no:

- Digital text corpora
- Annotated datasets
- Pos-tagged or syntactically parsed corpora
- No large-monolingual or parallel corpora
- No tagged corpora (POS, DEP, etc.)
- No morphological analysers
- No speech datasets (audio transcripts)
- …

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

NLP – CUSHITIC FAMILY

A language is Low-Resource if it has insufficient digitized text, annotated datasets, or NLP tools to train modern machine-learning models such as MT systems or large language models.

How the Cushitic Family is doing?

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

Language	NLP Resource Level	Corpora	MT	ASR	Pretrained Models
Somali	High	Yes (millions)	Yes	Yes	Yes
Oromo	Medium-High	Yes	Yes	Emerging	Yes
Saaho	Medium-Low	Yes (1M+ words, tagged)	No	No	Emerging
Afar	Low	Small	Yes	No	No
Beja	Very Low	Small (UD)	No	No	No
Sidama, Hadiyya, Rendille, Daasanach, etc.	Very Low	None/minimal	No	No	No

LANGUAGE

HO CORPUS

CUSHITIC FAMILY

DATA SOURCES

R INTERFACE

OD - EVALUATION

NCLUSIONS

NLP – CUSHITIC FAMILY

- Saaho Machine Translation Model is a novel machine translation system that utilizes a manually created parallel text dataset for training, combined with the Saaho Corpus multilingual (Saaho-English-Italian) and under construction monolingual database.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO DATA SOURCES

- The main source is Saho Corpus database which contains all the data collected by the previous scholars.
- New data provided by Asmara based Language Committee (Saho chapter) which is governmental institution.
- Scholars in the Diaspora of Saaho heritage.
- Communities around the globe (Saaho Diasporic Community)

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO DATA SOURCES

- The training data's strength comes from the diverse selection of source and target language texts, including common sentences from the Corpus build through the most frequently used words, and natural Saaho language examples contributed by native speakers, as well as idiomatic and culturally-rooted expressions.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO DATA SOURCES

- This helps address the challenge of small dataset sizes typically encountered in deep learning-based training for low-resourced languages.
- Very selective input data.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO DATA SOURCES

- USEFUL PHRASES: SAHO-ENGLISH. A collection of 200 sentences asked in Saho and translated into English collective from the Saho Community in London.
- Starting from the 500 most-frequently used words from Saho Corpus, constructed 500 sentences that use at least one of those words. Translated into English.
- Two books (short story and novel) translated from Saho to English.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO DATA SOURCES

- Concordances Saho – a list created by Moreno Vergari (2025) (non published). 145 words that had a frequency of > 50 in a 15.000 conjugate words corpus.
- Again got 145 sentences that use at least one of those words. Translated into English.
- Idiomatic expressions collected and translated by Ahmedsaad Omar.

THE LANGUAGE
SAAHO CORPUS
MT FOR CUSHITIC FAMILY
SAAHO DATA SOURCES
USER INTERFACE
SAMTRAMOD - EVALUATION
CONCLUSIONS

SAAHO DATA SOURCES

Finally, a collection 629 documents, including

- 260 Textbooks
- 136 Radion programs
- 13 Newsletter/magazine
- 10 Academic works
- 4 Dictionaries (word lists, etc.)
- 3 Poetry collection
- 2 Legal documents
- 2 Proverbs collections
- 196 Social Media, Others

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAAHO DATA SOURCES

- The main entrance for the English source language is KIT5K-Ver 1.0 from Hugging Face Community (a collection of 5 thousand sentences considered as standard frequently used contemporary English).

- The target, by the end of the year (2025) is to reach 15,000 sentences, parallel text Saaho - English selected accurately, with 10,000 from Saaho and 5,000 from English, translated bilingual.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

USER INTERFACE

- A private component for correcting data, editing and approving the parallel text.
- A public component for using for translations, bilingual for now English <=> Saaho.
- The public component also helps to collect data from the ordinary users to improve the parallel text.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

Welcome to the new Parallel Text management app. You can voluneteer to translate or check translation here

Searching the care in Saaho. 62 sentences identified. Show more verbose list.

Sentence ID	English	Saaho
K5000002158	Adam's roommate often leaves dirty dishes in the sink	*Adamlih caredde sigati mangole gedda, askok le kishshot nuwaa amaa nuwaa adde cakalishanidde xaba*
SC000005218	After hearing this, the old man felt disheartened and came back to his house	*Toye yoobbek sarrah, amaa naba xiyawti niya cideh ishi careh gaxe*
K5000002225	As soon as she entered the room, I rose to my feet and made a deep bow	*Ishi amaa caredde sayte waqhte akkanih, saf-erhxeyaako sirii ginah cunuune*
P0000008215	As soon as she entered the room, I rose to my feet and made a deep bow	*Ishi amaa cared orobte gul akkanih, xaf-erhxeyaako siriiginah ceege*
P0000008642	As soon as she entered the room, I rose to my feet and made a deep bow	*Ishi amaa caredde orobte geddaako, akkanih daw-erhxeh sirii ginah cunuune*
SC000005225	Both of them visited the house and found that the man died a day before, after waiting for a single reply from his daughter	*Usuk ishi barhäako wuli radde ilaalak yinemko, usun lammuk amaa care gufen maxko afti max, rabeh akah suge*
P0000005488	Both of them visited the house and found that the man died a day before, after waiting for a single reply from his daughter	*Usuk ishi barhäako inki radde ilaalaa yinemko, usun lammik amaa care gufen maxko afti max, rabeh akah suge*
P0000005598	Both of them visited the house and found that the man died a day before, after waiting for a single reply from his daughter	*Usuk ishi barhäako inki radde ilaalak yinemko, usun lammuk amaa care gufen maxko afti max, rabeh akah suge*
K5000000283	Dan and Linda returned home	*Daan kee Linda careh gaxen*
SC000006134	Do you have a house for rent?	*Atu kiraah care lito?*
SC000006135	Do you have a house for rent?	*Atu kiraa care litoho?*
P0000006268	Do you have a house for rent?	*Atin kiraa care litiiniho?*
P0000006269	Do you have a house for rent?	*Atin kiraah yakke care litiini?*
SC000005068	Do you have your own house?	*Atu ishi care lito?*

Add a new sentence | Total sentences: 7886 | Approved: 2472 | Checked: 934 | Edited: 36 | Not assingned: 4444 | My sentences: 6 | General report

Welcome to the new Parallel Text management app. You can voluneteer to translate or check translation here

Searching the

Sentence ID
K5000002158
SC000005218
K5000002225
P0000008215
P0000008642
SC000005225
P0000005488
P0000005598
K5000000283
SC000006134
SC000006135
P0000006268
P0000006269
SC000005068

The sentence: **K5000002158**: *Adam's roommate often leaves dirty dishes in the sink*

Adam's roommate often leaves dirty dishes in the sink

Sentence ID	2158
Unique sentence ID	K5000002158
English	Adam's roommate often leaves dirty dishes in the sink
Saaho	Adamlih caredde sigati mangole gedda, askok le kishshot nuwaa amaa nuwaa adde cakalishanidde xaba
Dialect	(•) Dialect neutral ( ) North dialect ( ) Central Dialect ( ) South Dialect
Notes	Updates on Kit 5K Ver 1.0 - Sep 2025. Translated by Suleyman Mohammed.
Status: approved	Save the translation Last modified 2025-11-19 14:10:02 ahmedsaad

Please edit only Saaho

Other information

Duplicates:

- Adamlih qorfadde sigati wuliili gedda, askok le kishshot nuwaa amaa nuwaa adde cakalishanidde xaba (Dialect: North dialect)
- Adamlih qorfadde marati yinemko, askok le kishshot nuwaa amaa nuwaa adde cakalishanidde xaba (Dialect: North dialect)

Add a new duplicate

Do you have your own house? Atu ishi care lite?

Add a new sentence | Total sentences: 7886 | Approved: 2472 | Checked: 934 | Edited: 36 | Not assingned: 4444 | My sentences: 6 | General report

Welcome to the new Parallel Text management app. You can voluneteer to translate or check translation here

Searching the

Sentence ID
K5000002158
SC000005218
K5000002225
P0000008215
P0000008642
SC000005225
P0000005488
P0000005598
K5000000283
SC000006134
SC000006135
P0000006268
P0000006269
SC000005068

The sentence: **K5000002158**: *Adam's roommate often leaves dirty dishes in the sink*

Adam's roommate often leaves dirty dishes in the sink

Sentence ID	2158
Unique sentence ID	K5000002158
English	Adam's roommate often leaves dirty dishes in the sink
Saaho	Adamlih caredde sigati mangole gedda, askok le kishshot nuwaa amaa nuwaa adde cakalishanidde xaba
Dialect	◉ Dialect neutral ○ North dialect ○ Central Dialect ○ South Dialect
Notes	Updates on Kit 5K Ver 1.0 - Sep 2025. Translated by Suleyman Mohammed.
Status: approved	Save the translation Last modified 2025-11-19 14:1[illegible]02 ahmedsaad

Please edit only Saaho

Other information

Duplicates:

- Adamlih qorfadde sigati wuliili gedda, askok le kishshot nuwaa amaa nuwaa adde cakalishanidde xaba (Dialect: North dialect)
- Adamlih qorfadde marati yinemko, askok le kishshot nuwaa amaa nuwaa adde cakalishanidde xaba (Dialect: North dialect)

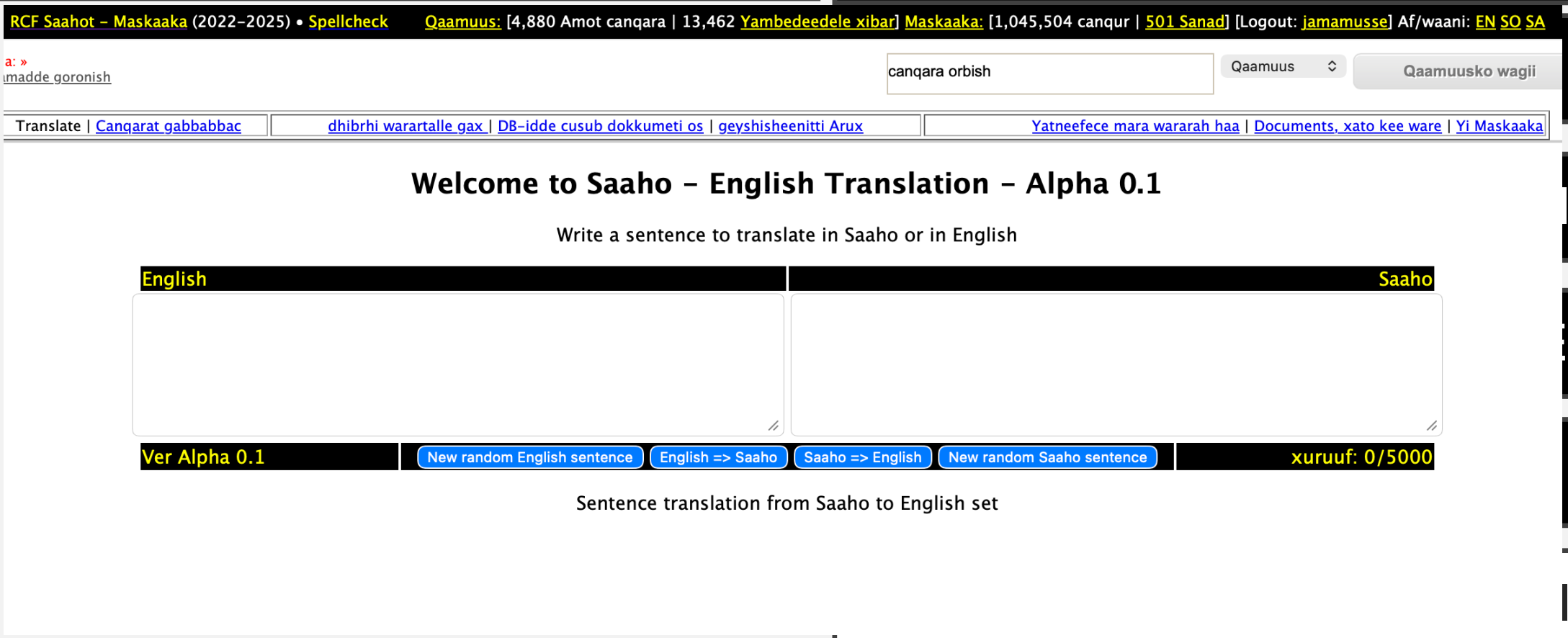
Add a new duplicate

Do you have your own house? Atu ishi care lite?

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

USER INTERFACE

THE LANGUAGE

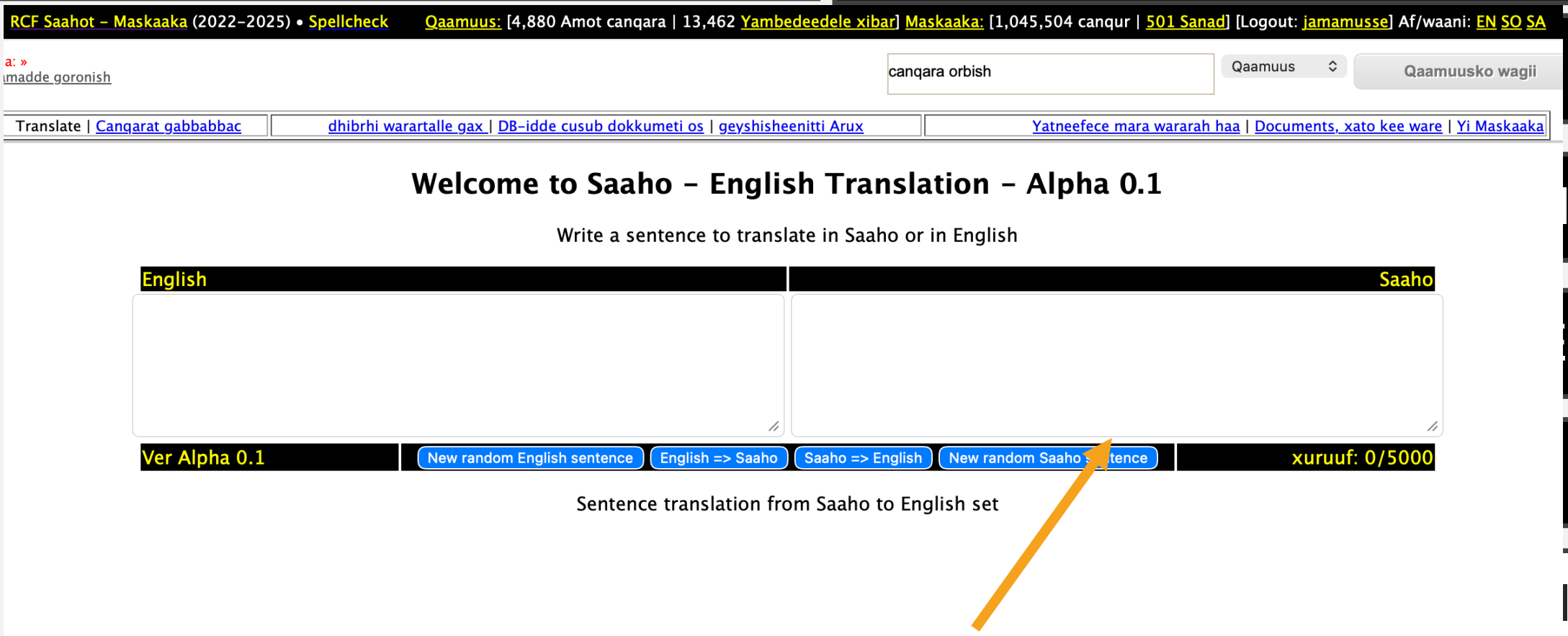


CONCLUSIONS

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

USER INTERFACE

THE LANGUAGE

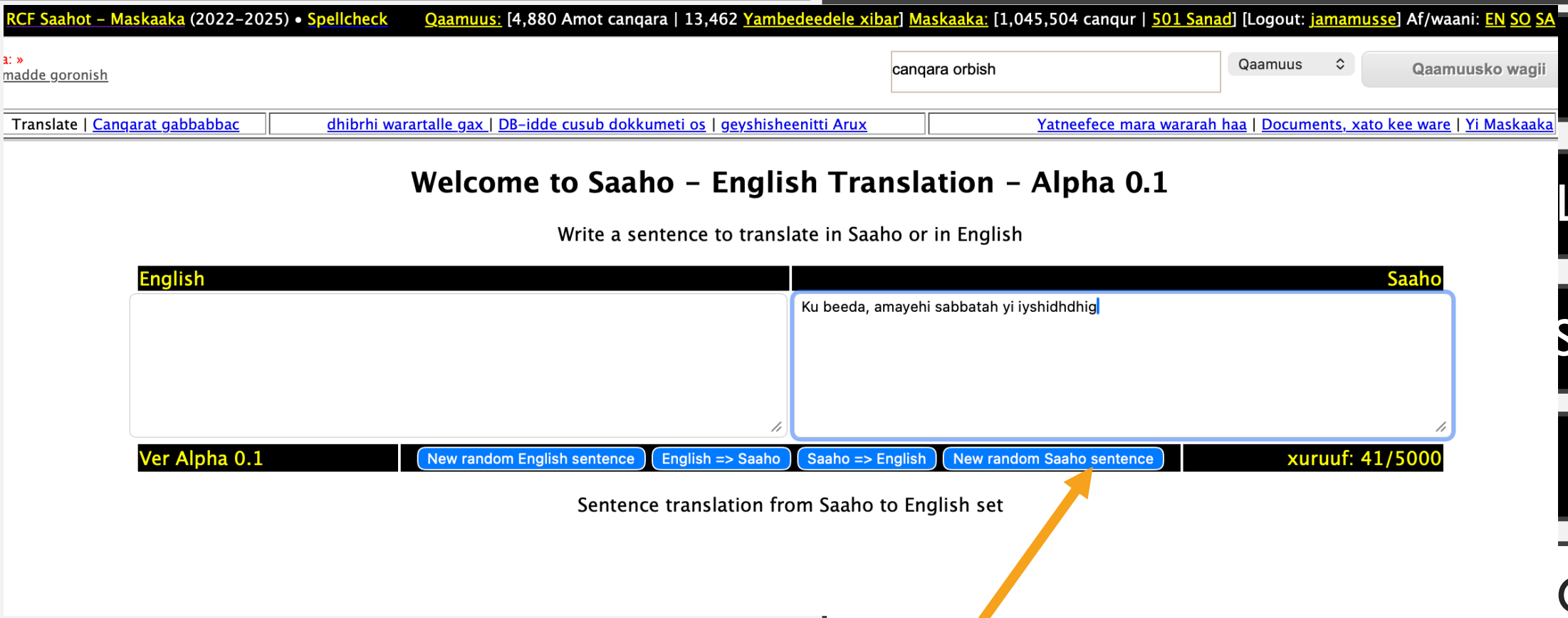


CONCLUSIONS

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

USER INTERFACE

THE LANGUAGE

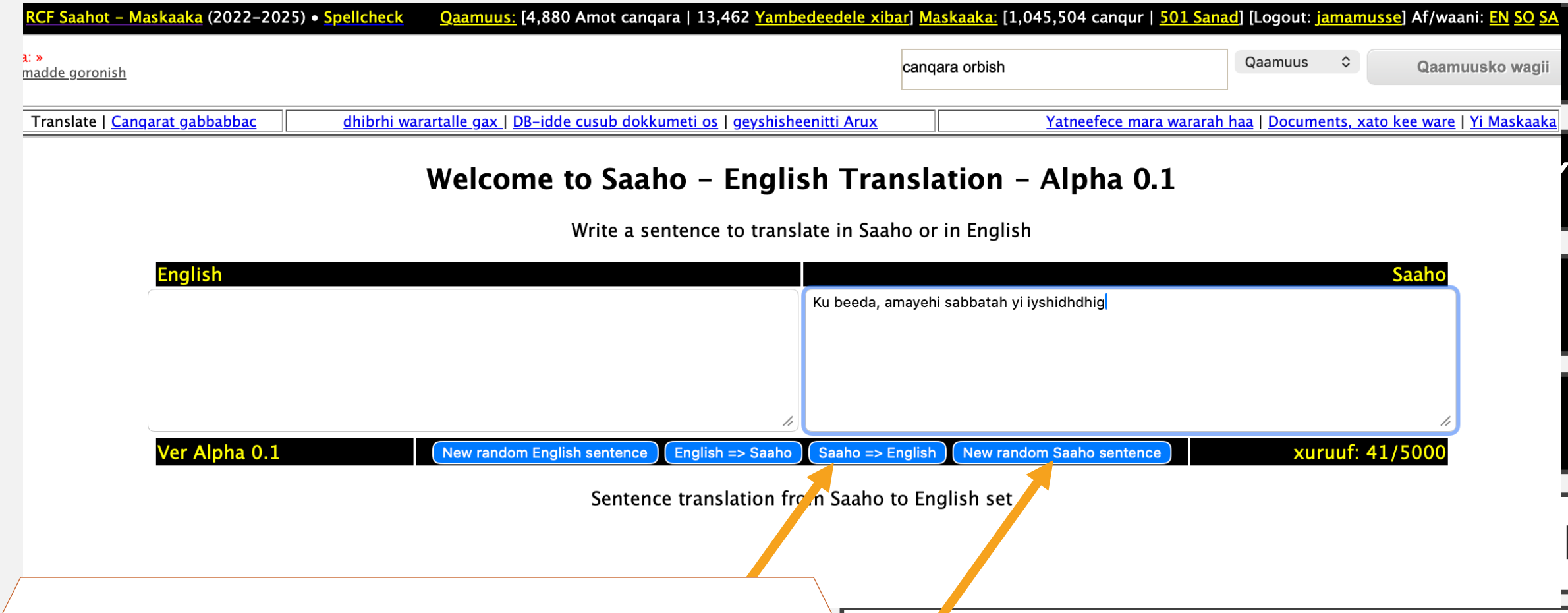


CONCLUSIONS

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

USER INTERFACE

THE LANGUAGE



CONCLUSIONS

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

USER INTERFACE

THE LANGUAGE

SAMPLING METHOD - EVALUATION

CONCLUSIONS

SAMTRAMOD

- Saaho Machine Translation Model is a novel machine translation system that trained with this manually created parallel text.
- It challenges the narrative that Modules created for major Western Languages where you have billions and billions of datasets, should/could be also used for Low-Resourced Languages.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

SAMTRAMOD

Data input

The PILE - SAMTRAMOD

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

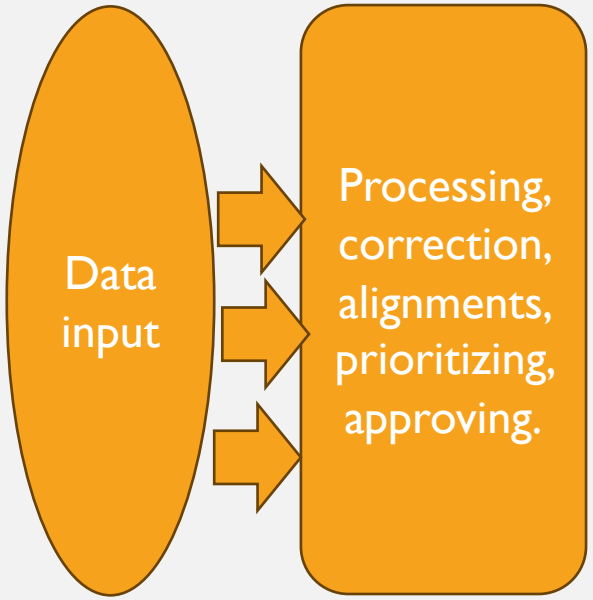
SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAMTRAMOD



THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

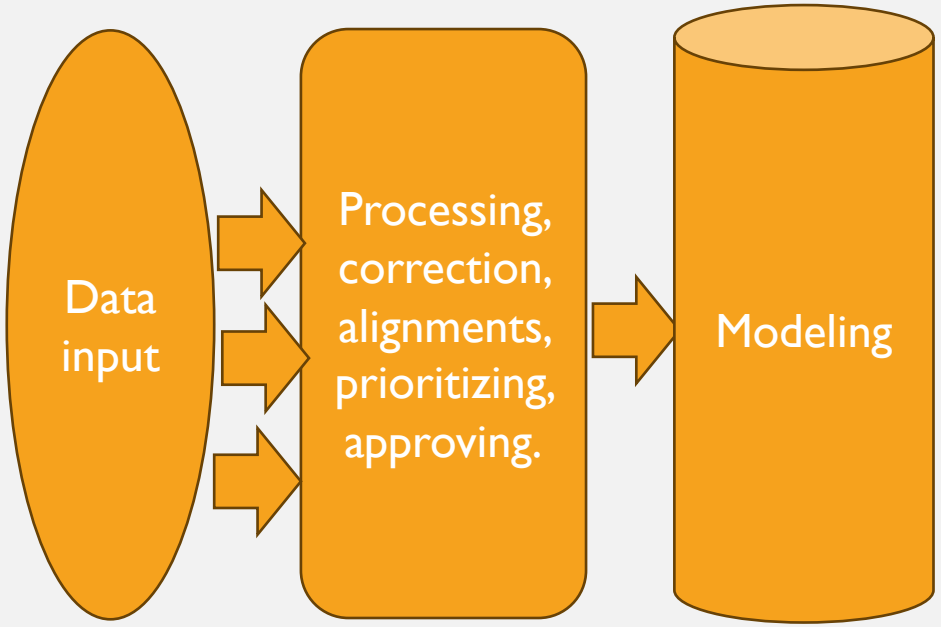
USER INTERFACE

SAMTRAMOD - EVALUATION

The PILE - SAMTRAMOD

CONCLUSIONS

SAMTRAMOD



THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

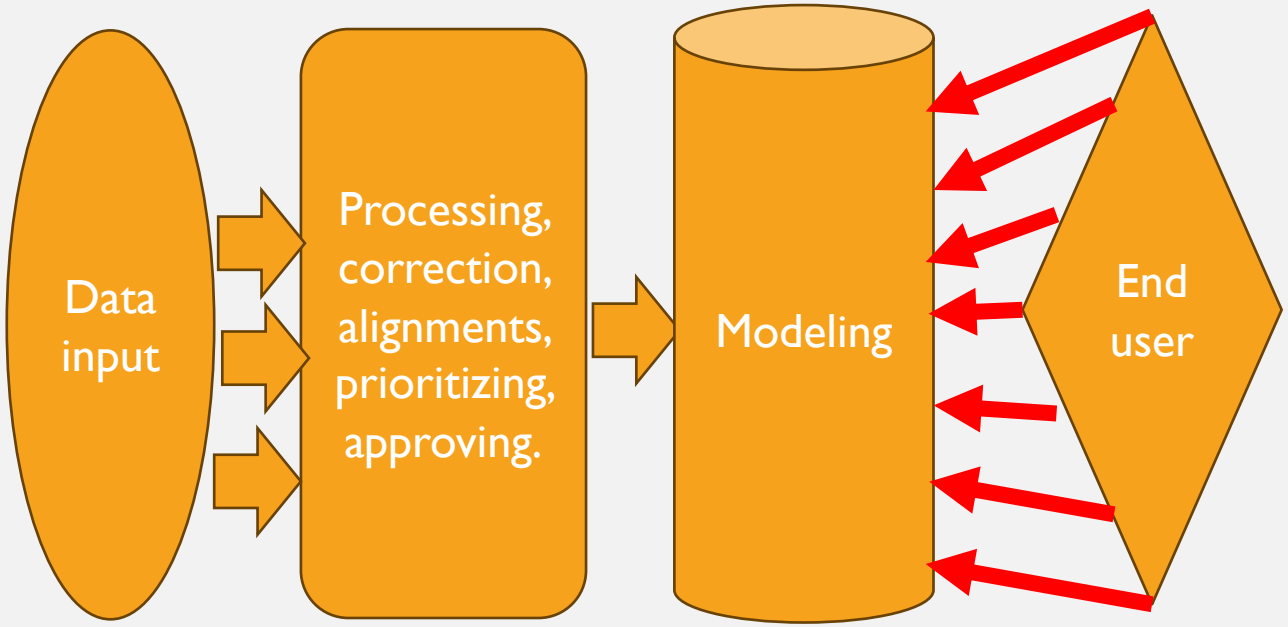
USER INTERFACE

SAMTRAMOD - EVALUATION

The PILE - SAMTRAMOD

CONCLUSIONS

SAMTRAMOD



End user could be an APP, Human, another platform, another model.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAMTRAMOD

Automatic Alignment.

- All sources and target pair has been passed to *memoryln* automatic parallel corpus.
- A candidate generation step produces candidate pairs, and cleans one more time.
- The system is similar to Junczys-Dowmunt(2018) where inverse translations models are used to weighted average aggregated over a set of candidate pairs.

THE LANGUAGE
SAAHO CORPUS
MT FOR CUSHITIC FAMILY
SAAHO DATA SOURCES
USER INTERFACE
SAMTRAMOD - EVALUATION
CONCLUSIONS

SAMTRAMOD

Evaluation

- The author met in Nairobi, Addis Ababa, Sharjah, Hawassa, Djibouti and London (2025), and made human evaluation of the translations.

- Sadly, there is no state-of-the-art commercial systems (Google Translate, Microsoft Translator, etc.) for Saaho.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAMTRAMOD

- The results are promising but not as good as I wished because of small data we engaged for this translations.
- So, what are the next steps?
- For sure continue working on a culturally sensitive data, but also to try parallel act.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

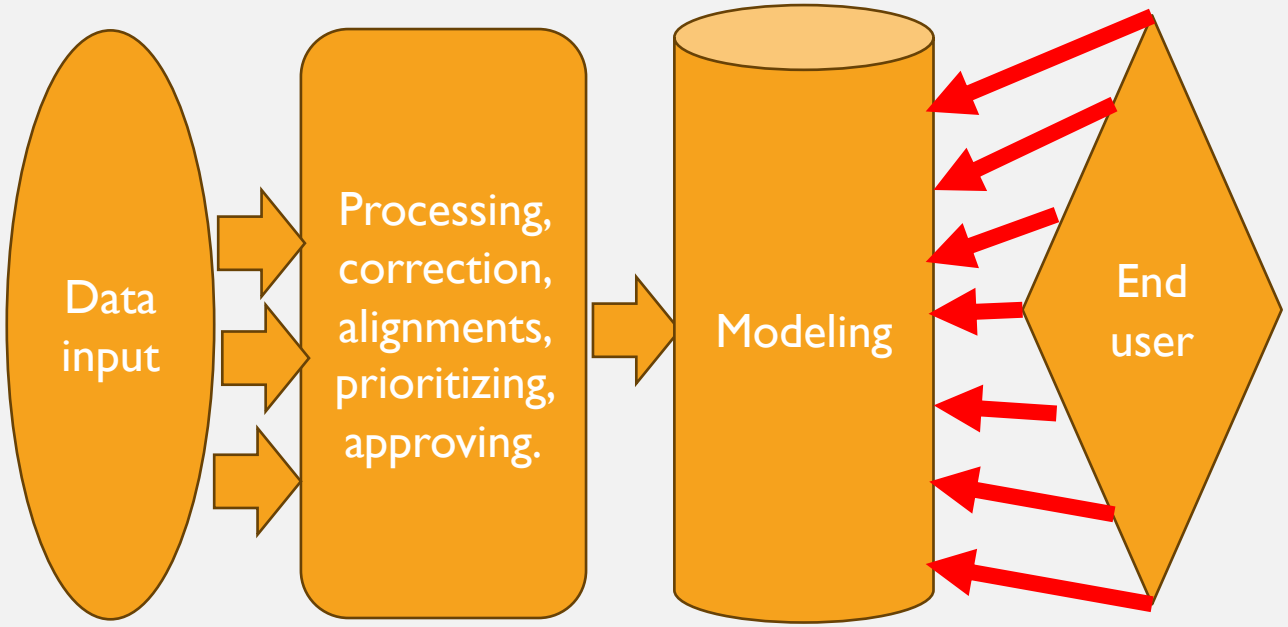
SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAMTRAMOD



What happen if I use Large Language models (LLM) and AI?

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

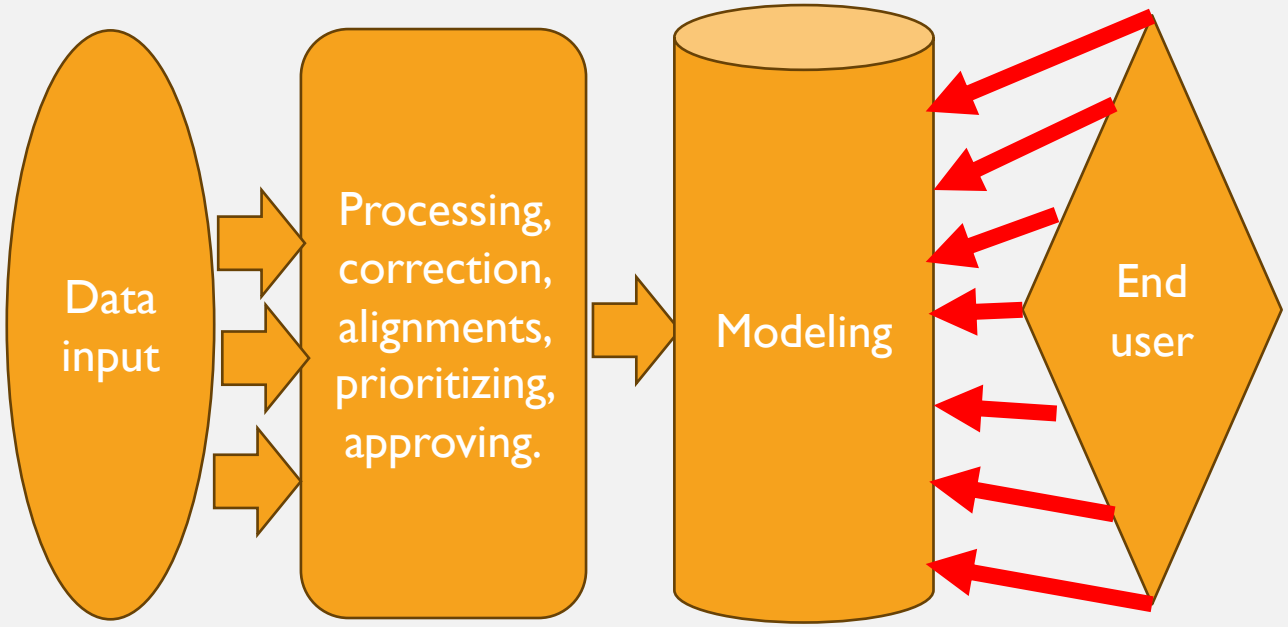
SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAMTRAMOD



GPT4, DeepSeek-R1, Llama, Mistral AI, BERT, Grok, OpenAI, ChatGPT, …

What happen if I use Large Language models (LLM) and AI?

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

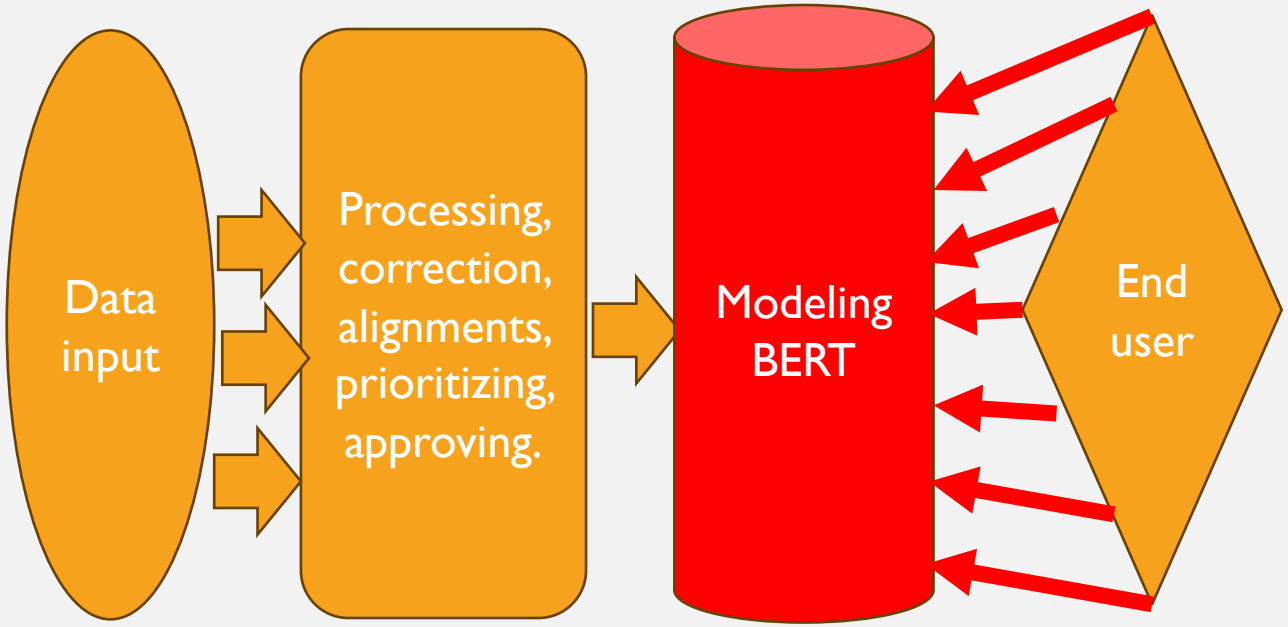
SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

SAMTRAMOD



What happen if I use Large Language models (LLM) and AI?

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

CONCLUSIONS

Broader impact (1)

- Today we have 7.1M articles in English in Wikipedia and over 18 billion words.
- Unfortunately, many people speaking in other languages (like Saaho) might not have access to that knowledge.
- SAMTRAMOD will help not only basic needs of translations, but can translate other major Large data.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

CONCLUSIONS

Broader local impact (II)

- Somali-Saho translator will be quite straight forward because of the common structure of the corpus, and the module being build specifically in this prospect.

- Cushitic – Cushitic translations will be feasible without passing through English or other third language step.

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

ACKNOWLEDGMENT

Thanks to

- Saaho Speaking Community in Eritrea and and the Diaspora
- Saaho Corpus Team (in London and Asmara – Language Group)
- Saaho scholars, Ahmedsaad Omar and Ahmed Suleiman in London.
- Giorgio Banti and Moreno Vergari for their constant contribution with ideas and materials

THE LANGUAGE

SAAHO CORPUS

MT FOR CUSHITIC FAMILY

SAAHO DATA SOURCES

USER INTERFACE

SAMTRAMOD - EVALUATION

CONCLUSIONS

Saahot Afhi Maskaaka

Canqurko 1 miliyoonko bukalle takkem yemweesebe | 440 takkem dokkumentitko barhsimeh tuumaarime | amot canqurko 4700 bukalle takkem geytime yimfillice canqurko 20 shixtiya takkem yewce | 4500ko bukalle yakke canqur 2 luqhah yunturjume

Welcome to Saaho – English Translation – Alpha 0.1

Write a sentence to translate in Saaho or in English

English	Saaho
Thank you!	Gaab yoh isho!

Gaab yoh isho!

Ver Alpha 0.1 | New random English sentence | English => Saaho | Saaho => English | New random Saaho sentence | xuruuf: 0/5000

Sentence translation from Saaho to English set

Gor kin Angollal

- Redsea Cultural Foundation, Hargeysa, Somaliland
- Oriental University of Naples, Naples, Italy
- Hargeysa Cultural Centre, Hargeysa, Somaliland

Cawwatoyshinnan aki angollalah

- Saahot Makaado, London, UK
- Ethnorema Bolzano, Italy
- Saho Speaking Community in London, London, UK

Dhayih tinthibbicem

- Saho Corpus: Semi-automation of Verb Conjugation in Saho: Verbs Class I, Jama Musse Jama, 2022
- Saho Corpus: Foundations and the Current State, Jama Musse Jama, 2021

Xato kee ware

- Yatneefecetiyahi santo / User Manual
- Most frequently asked questions
- Subscribe to our newsletter
- Follow us on Facebook and Twitter
- NEW Upload files to the corpus

REDSEA CULTURAL FOUNDATION
www.redsea-online.org

RCF Saahot – Maskaaka (2022–2025) • Spellcheck Qaamuus: [4,880 Amot canqara | 13,462 Yambedeedele xibar] Maskaaka: [1,045,504 canqur | 501 Sanad] Af/waani: EN SO SA

Saahot Afhi Maskaaka

Canqurko 1 miliyoonko bukalle takkem yemweesebe | 440 takkem dokkumentitko barhsimeh tuumaarime | amot canqurko 4700 bukalle takkem geytime yimfillice canqurko 20 shixtiya takkem yewce | 4500ko bukalle yakke canqur 2 luqhah yunturjume

tarjama: »
tarjamadde goronish

canqara orbish Qaamuus Qaamuusko wagii

Translate | Canqarat gabbabbac | dhibrhi warartalle gax | DB-idde cusub dokkumeti os | geyshisheenitti Arux | Yatneefece tiyahi Addoodim eymemec | Documents, xato kee ware | Yi Maskaaka

Welcome to Saaho – English Translation – Alpha 0.1

Write a sentence to translate in Saaho or in English

English	Saaho
Thank you!	Gaab yoh isho!

Thank you! Gaab yoh isho!

Ver Alpha 0.1 | New random English sentence | English => Saaho | Saaho => English | New random Saaho sentence | xuruuf: 0/5000

Sentence translation from Saaho to English set

Gor kin Angollal

- Redsea Cultural Foundation, Hargeysa, Somaliland
- Oriental University of Naples, Naples, Italy
- Hargeysa Cultural Centre, Hargeysa, Somaliland

Cawwatoyshinnan aki angollalah

- Saahot Makaado, London, UK
- Ethnorema Bolzano, Italy
- Saho Speaking Community in London, London, UK

Dhayih tinthibbicem

- Saho Corpus: Semi-automation of Verb Conjugation in Saho: Verbs Class I, Jama Musse Jama, 2022
- Saho Corpus: Foundations and the Current State, Jama Musse Jama, 2021

Xato kee ware

- Yatneefecetiyahi santo / User Manual
- Most frequently asked questions
- Subscribe to our newsletter
- Follow us on Facebook and Twitter
- NEW Upload files to the corpus

GE
US
FAMILY
URCES
ACE
LUATION
NS

THE SAAHO LANGUAGE

SAMTRAMOD

REFERENCES

MORIN, Didier. 1999. *Le texte légitime. Pratiques littéraires orales traditionnelles en Afrique du nord-est*. Louvain: Peeters.

MORIN, Didier. 2006. 'Asaurta. In Steffen Wenig (ed.), *Im kaiserlichen Auftrag - Die Deutsche Axum Expedition 1906 unter Enno Littmann*, vol. 1,77-82. Aichwald: Linden Soft.

MORIN, Didier. 2015. 'Afar-Saaho dialectology: a methodology. *Ethnorêma* 11.19-54.

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

ZABORSKI, Andrzej. 1987. 'Afar-Saho linguistics - an overview. In Hans G. Mukarovsky (ed.), Leo Reinisch - Werk und Erbe, 85-95. Wien: Österreichische Akademie der Wissenschaften.

VER 1.0

References

BANTI, Giorgio, and Moreno VERGARI (2005) "A Sketch of Saho Grammar", *Journal of Eritrean Studies* 1-2: 100-131.

BANTI, Giorgio, and Moreno VERGARI (2023) "Saaho". In MEYER, Ronny, Bedilu WAKJIRA, and Zelealem LEYEWM (eds.), *The Oxford Handbook of Ethiopian Languages*. Pp. 294-319. Oxford: Oxford University Press.

DE MARNEFFE, Marie-Catherine, Christopher MANNING, Joakim NIVRE, and Daniel ZEMAN (2021) "Universal Dependencies", *Computational Linguistics,* 47(2): 255-308.

DI MANNO, Andrea and MILIZIA, Paolo (2024) "Some preliminary observations on the preparation of a Saho treebank I", *Ethnorêma* 20: 29-40.

HAYWARD, Richard J. 1979. Some inferences from an irregular imperative form in Saho. *Israel Oriental Studies* 9.245-257.

HAYWARD, Richard J. 1983. Some aspects of the phonology of ultimate vowels in Saho-'Afar. In Stanislav Segert & Andras J.E. Bodrogligeti (eds.), *Ethiopian studies dedicated to Wolf Leslau on the occasion of his seventy-fifth birthday November 14th, 1981 by friends and colleagues*, 221-231. Wiesbaden: Otto Harrassowitz.

HAYWARD, Richard J. 1997. External sandhi in the Saho noun phrase. *Afrikanistische Arbeitspapiere* 50.53-80.

HAYWARD, Richard J. & Martin Orwin. 1991. The prefix conjugation in Qafar-Saho: the survival and revival of a paradigm - part I. *African Languages and Cultures* 4(2). 157-176.

JAMA MUSSE JAMA (2025) "Saho Corpus: Annotating gender polarity and other morphological features in Saaho nouns, *Rassegna di Studi Etiopici* 34.127-168.

JAMA MUSSE JAMA (2021) "Saho Corpus: Foundations and the Current State", Available at www.sahocorpus.com *(accessed 15 October 2024)*.

JAMA MUSSE JAMA (2022) "Saho Corpus: Semi-automation of Verb Conjugation in Saho: Verbs Class I", *Ethnorêma* 18: 69-87.

JAMA MUSSE JAMA (2016) "A Syntactically Annotated Corpus of Somali Literature". PhD thesis [unpublished]. University of Naples "L'Orientale". See www.somalicorpus.com

JAMA MUSSE JAMA and TIRSIT YETBAREK (2024) "Extension of Somali Corpus to Omo-Tana branch of the Cushitic Languages", *Dhaxalreeb*, 20 (1).

LAMBERTI, Marcello. 1990. Some word order principles of Saho-Afar. *Rassegna di Studi Etiopici* 34.127-168.

MORIN, Didier. 1995. *'Des paroles douces comme la soie.' Introduction aux contes dans l'aire couchitique (bedja, afar, saho, somali)*. Louvain: Peeters.