III SHO‘BA
NLPDA TIL VA NUTQ TAHLILI
(LINGVISTIK ANALIZATOR: MORFOLOGIK, SINTAKTIK VA SEMANTIK TAHLIL, NUTQ ANALIZI VA SINTEZI)

MAKING USEFUL TOOLS FOR LANGUAGE COMMUNITIES WITHOUT BIG CORPORA: A VERSATILE MORPHOLOGICAL ANALYSER AND SPELLCHECKER FOR SONGHAY BASED ON FREE SOFTWARE

Loïc Boizou
PhD, researcher
Department of Digital Humanities
University of Cologne.

Mohomodou Houssouba
PhD, researcher
Centre for African Studies
University of Basel.

Abstract: The current excitement over data-driven artificial intelligence makes headlines across the world, including in Africa. Accordingly, AI is supposed to enable an exponential growth in digital resources and level the playing field for less-resourced languages. However, this enthusiasm largely eludes the correlation between AI capacity and the availability of massive and well-curated datasets. In this regard, for the most part, African languages do not fulfill the basic requirements to participate in the new “revolution”. The Songhay language is a case in point, with slow and continuous work done over two decades to develop even a basic tool like a spellchecker. Still, even limited experience with localization and building corpora helps toward developing more advanced tools and content over time.

Keywords: free software - spellchecker - morphological tagging - morphology - Songhay language - under-resourced languages

Introduction

While big data and artificial intelligence make the headlines, the increasing digital rift is getting less attention in the media landscape. Only a small number of the world languages, roughly a big hundred, have access to the digital Graal while thousands are still practically or even totally left outside. Even for the languages that entered the digital space, there is a wide range of concrete situations, from AI-supported languages to languages with rough elementary resources. The situation is especially dire for African languages: practically no endogenous African language can be considered as well supported in the digital space. Furthermore, since cutting-edge technologies are highly data-intensive, most world languages, which are hardly written, let alone standardized, cannot really benefit from the development of these technologies. The big divide risks to widen further, unless a boom in speech-to-text technology can somehow mitigate the situation.

This article presents a Hunspell spellchecker and tagger designed for the Songhay language, a language spoken by several million speakers across borders, in Niger, Mali, Burkina

Faso, and Benin. Even if spellcheckers are a relatively old technology, they are still relevant for everyday users as a convenient help to avoid misspellings. As such, spellcheckers constitute a useful language-related technology, as well as some more complex tools such as machine translation systems, optical character recognition, speech-to-text systems or conversational agents. Traditional spellcheckers are quite easy to develop using standard mature tools such as Hunspell, on the basis of a lexicon and, if needed, a morphological model. Although not as advanced as predictive typing, which takes into account the context (the previous words), it avoids the need for significant corpora, which is still a missing resource for a huge number of languages.

The present contribution brings to the fore the weak integration of the African languages in the digital domain, to explain the rationale for developing a Songhay Hunspell module. It offers a general overview of the design of the Songhay morphological model, shows the two main usage cases, as spellchecker and morphological tagger and, finally, opens up perspectives for further enhancements of the tools for the language.

1. African languages in the digital space

As the internet opened on a global scale in 1996, most interfaces and contents were almost exclusively in English. Programmers had been mostly trained in English, and many wouldn't be able to work in their native languages, so even major international languages were sidelined at the outset. In most African countries, literacy and knowledge production were realized in European languages (English, French, Portuguese and Spanish) and Arabic. In many ways, the emergence of English as the nearly exclusive idiom of the new digital age prompted a feverish debate over cultural diversity, social equality and political participation. Paradoxically, a postcolonial framework like the Francophonie would become a vocal advocate for inclusiveness. Thus, African members of the Latin Union and Arab League played a major role in pushing the diversity agenda. Some participated in all three networks and related consortiums. The stakes were deemed high and existentially significant to major languages, not to mention the less-resourced ones spoken across Africa.

The International Francophone Network for Language Management (RIFAL) was a key instrument in coordinating multi-pronged initiatives at official and associative levels across countries. To be sure, while the institutional focus was on creating a French-speaking ecosystem in cyberspace, there were competing visions within the Francophonie. In many countries, French as an official language, though predominant in official business, was and is still spoken by only a small urban elite. The vast majority of the population communicates through local languages and creolized forms of French. Their inclusion into the digital age would require creating various kinds of resources on a sustainable basis. This imperative pushed language activists to aim beyond the OIF framework, by learning from different experiences and adapting the most accessible ones to their native environments.

It is worth noting that earlier projects tried, often in isolation, to produce tools for specific purposes. For instance, the Kamusi Project started in 1994 as a "living" Swahili dictionary, then evolved to become a multilingual platform linking over 40 languages across continents. Many initiators of similar single projects would come in contact once the internet enabled broad and continuous networking across campuses and regions. This was a pioneer era in which individuals and volunteer groups actually had the biggest impact in putting African languages on the map. By 2000, several networks emerged at the forefront. The PanAfrican Localization project (PAL)

quickly federated disparate efforts to translate software interfaces into African languages. It was a daunting task at a time of low-speed and costly dial-up connections. The African Languages Technology Initiative (Alt-I) was another effort to remove technical barriers by sharing know-how. Bisharat greatly contributed to aiming at the larger population, focusing on the development needs of rural communities – agriculture, animal and human health, or small trade [Osborn, 2010].

This loose coalition of actors also played a significant role in advocacy and lobbying work to effect policy change. After 2005, it coalesced into a younger and more structured platform, the African Network for Localization (ANLoc) based in South Africa, with a strong connection with Arabize in Egypt and technically savvy teams in Kenya, Ghana, Uganda, among others. With strong support by the Canadian International Development Research Centre (IDRC), the group carried out a coherent agenda to engage African languages with even a handful of volunteers willing to devote time and energy to the cause [Bailey, 2012].

The Songhay localization project, which started as a one-person initiative in 2004, really took off only in 2007, when a technically savvy volunteer joined and started to push translated user interfaces through the release pipelines. It joined the ANLoc network in 2009 and by 2012 registered a dozen released localizations, including Mozilla Firefox.

Apart from localization, access to language-related tools also remains scarce. Based on an overview considering two important tools for everyday users[15], spellcheckers and machine translation systems, the number of supported African languages is around 20 or sometimes even less.

Machine Translation systems for African languages

Google Translate + Facebook	Afrikaans, Amharic, Ganda, Hausa, Igbo, Lingala, Malagasy, Northern Sotho, Oromo, Somali, Swahili, Xhosa, Yoruba, Zulu
Google Translate only	Akan, Bambara, Chewa, Ewe, Kinyarwanda, Shona, Southern Sotho, Tigrinya, Tsonga
Facebook only	Pulaar-Fulfulde, Tswana, Swazi, Wolof

Hunspell modules for African languages (with number of words in the lexicon)[16]

Afrikaans (188,366), Akan (5007, full form), Amharic (182,552, full form), Bambara (26,012), Kituba (2171), Lingala (5194, some grammatical annotations), Malagasy (16,743, full form), Moore (105,696, full form), Sango (13,017), Tswana (61,598, full form), Shona (37,287), Swahili (Kenyan 48,317, full form, Tanzanian 67,900), Zulu (108,900, some grammatical annotations)[17]

[15] As far as virtual keyboards are concerned (also an important core technology), the number of Gboard configurations is far higher, but the predictive typing is often provided with almost no initial data (e.g. for Songhay).
[16] Theses modules were found in three locations: https://github.com/elastic/hunspell/tree/master/dicts, https://github.com/wooorm/dictionaries/tree/main/dictionaries, https://extensions.libreoffice.org/?q=Spellchecker&action_doExtensionSearch=Search
[17] There is also the very beginning of a Kinyarwanda Hunspell, but with 62 words only.

The wide scope of the size of the lexicons reveals the very diverse level of maturity of these Hunspell modules, yet the number of words in the lexicon alone can be partly misleading: for inflected languages, a relatively small lexicon with productive affixal patterns can be as expansive as a larger full-form lexicon. In certain mentioned Hunspell modules, e.g. Akan or Amharic, some patterns are defined in the affix files but are not used in the lexicon. No dictionary provides the structured set of grammar features (although the Lingala and the Zulu modules do include some pieces of grammatical information) that would allow the system to operate not only as a spellchecker but also as a morphological tagger.

The scarcity of the resources for the African languages, which is illustrated with both examples, is mainly the consequence of a sharp linguistic imbalance inherited from the colonial situation which left English, French and Portuguese (also Spanish to a far lesser extent) as strongly dominant languages [Ebongue and Hurst, 2017, more specifically Nakayiza and Ssentanda 2017, Kamwangamalu, 2019] as well as a pattern of state unilingualism [Ekkehard Wolff, 2017]. In some countries, Arabic (outside its African core territories) too imposes itself over local languages through religious prestige or official language policymaking. In almost all African countries, these mentioned languages are *de jure* or *de facto*[18] state languages (Ethiopia being an exception). When African languages are official or co-official, their role is most frequently purely symbolic (except for Somali and Swahili). Thus, exogenous languages dominate domains such as administration, education, big business, often religion, while African languages remain mainly vernacular or, at best, spoken vehicular and initial literacy mediums in primary schools. As a secondary consequence, there is a lack of available written accessible material for IA-based data-hungry technologies. Finally, given the administrative underuse of the local languages and the economic weakness of most African countries, the financial gains that could be expected from investing in resources for African languages are meager. The outcome might improve with the development of African middle classes, but only under the condition that this movement does not accelerate the relegation of endogenous languages. In such a grim context, each concrete tool for under-resourced African languages can have a slight positive outcome.

2.1 Rationale for Hunspell

Hunspell[19] is the most common tool for designing spellcheckers. It is free, mature and stable and well supported especially in the context of free software like Linux, which should be an important aspect to have in mind when fundings is scarce. It is the standard spellchecker for LibreOffice tools, for most browsers (Firefox, Chrome, Opera) and can be combined with the textual database ElasticSearch or used through Python[20] or R[21] libraries. The Hunspell files can be provided directly, but LibreOffice allows packing them and recording them as an extension on their extension bank. All these characteristics ensure Hunspell systems a good level of availability for any language community.

[18] Mali is a striking example of this situation. According to the Government Gazette of Mali (Décret N°2023-0401/PT-RM du 22 juillet 2023 portant promulgation de la constitution), French is not an official language, but is the working language. Nevertheless, as an obvious illustration of the *de facto* official status of French, this Government Gazette is published in French only.

[19] http://hunspell.github.io/

[20] https://pypi.org/project/hunspell/

[21] https://cran.r-project.org/web/packages/hunspell/vignettes/intro.html

Hunspell systems are a combination of lexicon and affix files encoded as Unicode plain texts. Therefore, Hunspell spellcheckers can be designed with simple unicode editors. As a drawback, it is easy to corrupt the data while developing or correcting Hunspell models, but this issue does not affect the end users.

Hunspell has a second potential benefit: if the model is properly designed, it can be used as a morphological tagger and lemmatizer. Morphological analysers are not a resource for everyday users but they are a traditional tool to provide annotated language data for linguists. Using Hunspell for both usage cases is an efficient way to avoid work reduplication for under-resourced languages. As for spellchecking, this task is carried out over isolated wordforms, without contextual disambiguation. It is an obvious limitation, but it a also what allows the tool to be designed without available corpus[22].

Previous Resources

Two pre-existing resources allowed developing the Songhay Hunspell Module very quickly.

The core element of Hunspell systems is a sufficiently comprehensive and cohesive lexicon. The development of the Songhay Hunspell module relies on previous work by the authors of the present contribution, as well as Abdoulbaki Seydou Cissé, who is the main developer of the website songhay.org, and the leading Songhay lexicographer Youssouf Mohamed Haïdara [Haïdara, 2010].

A draft morphological model was also ready. It was developed while designing a previous analyser for Songhay in Python, but a decision was made to switch to a widely available standard solution such as Hunspell in order to increase its availability. The initial model based on a traditional grammar (mainly the short introduction in Prost, 1956) was improved by generating all nominal inflected forms for the given lexicon and by checking the output manually. In this initial model, the nominal inflectional component was only developed for substantives, excluding adjectives, numerals and pronouns. The new enhanced model is described in the next section.

3. The designed Songhay Morphological Model

The present Hunspell module is designed for Mali's standardized Songhay, which is based on the Gao dialect, with minor adaptations needed to take into account localized specificities (e.g. -ante / -anta). Standardization is necessary to make a coherent spellchecker: allowing clearly different dialectal variations in the same text would risk enabling very artificial text and make the model far more complex, thus compromising the maintenance of the lexicon.

Hunspell has some limitations in the number of generation steps that impact the design for agglutinative languages[23] or, even worse, for polysynthetic languages. For Songhay, which possesses a relatively simple inflexional morphology and relatively stable stems, Hunspell fits perfectly.

As already mentioned, the model was designed in order to be usable not only as a spellchecker but also as a morphological analyser. For that purpose, the two most important aspects are to link all suppletive forms to the lemma through the Hunspell tag *st:* and to provide a full set of morphological tags for each form (either stem or derived).

[22] But Hunspell can also be a starting point in this direction [Pirininen and Lindén, 2010].
[23] It would require long compound endings, e.g. number + possession + case.

Parts of Speech (number of words)	Categories
noun (6069 common + 21 suppletive forms, 11 proper) verb (1925) adjective (201) adverb (75) interjection (56) numeral (24 + 9 suppletive forms) pronoun (15 + 11 suppletive forms) auxiliary (11 + 5 suppletive forms) preposition (11) postposition (10) conjunction (6) determiner (4) particles/adverbs[24] (4)	negative (aux.) definite/indefinite (noun, adj, num + rare pron) singular/plural (noun, adj, num, pron) person 1, person 2, person 3 (pron) emphatic (pron) complement (pron) proper (noun) ordinal (num) transitive (verb) infinitive (verb) participle (verb)

The categories include different levels of information, besides classical morphological categories (number, definiteness, polarity, person), there are single tags to mark some particular features, for propers nouns, emphatic and complement pronouns and transitive (or factitive) verbs. There are also tags to mark infinitives and participles as well as ordinal adjectives derived from numerals.

The lexicon contains 8468 words, but the real number of lexical units must omit the 46 suppletive forms; that is, 8422 lemmas. The remarkably low number of adjectives is related to a feature of the Songhay language whereby a significant part of words that would be semantically similar to Indo-European adjectives are verbs e.g. *faala* 'to be easy, to be cheap'.[25]

The lexical coverage is extended by the systematic derivation patterns: verbal adjectives / participles, verbal nouns / infinitives and ordinal adjectives. Given the affixal classes (14 suffix classes and 1 prefix class), the approximate number of word forms is about 30,000 nouns, 1000 adj, 20,000 verb forms, other about 300, probably about 55,000 forms in total.

The Hunspell model includes four rewriting rules, that allow adding the caron to some consonants before the letter *i* and *e* when missing: *si* → *ši*, *zi* → *ži*, *se* → *še*, *ze* → *že*. Usage and even norm are hesitant since in *s* and *z* (usually [s], [z]) are regularly palatalized before i and e as [ʃ] and [ʒ] in the Gao standard pronunciation, making the marking redundant (but needed in a strict phonographic approach) and slightly altering the stem in some cases (*bolši* / bolsi 'to boast' → *bol**s**andi*).

3.1 The Nominal Component

Songhay nominals have regularly four different forms, e.g. the word *gallu* 'town':

- gallu (singular indefinite)
- galloo (singular definite)

[24] No strong decision was made for these words.

[25] Languages with verb-like adjectives [Stassen, 2013] are not a rarity, Japanese being a widely cited example [Baker, 2003].

- galley (plural indefinite)
- galluyan (plural definite)

The Hunspell component describes nominals in three different ways.

3.1.1 Productive types

For usual nominals, the three suffixed forms are generated from the base form (singular indefinite). Songhay has three main models (the plural definite has only one form and therefore has no influence on the paradigms):

Songhay nominal endings

	paradigm 1	**paradigm 2**	**paradigm 3**
Singular definite	-oo	-aa	-aa
Plural indefinite	-ey	-ey	-awey
number of base words (excluding words with several combinations)	3417	756	2057

Some words show a certain variation. There are 55 words that can appear as 2nd or 3rd paradigms, two words of 1st or 2nd paradigms, and three words from the 3rd paradigm that can also appear with the *-oo* ending. As it is obvious from the previous table, the second model can be described as a mixed one, with singular similar to the third group and plural similar to the first group. Therefore, the Hunspell model was built about two singular suffixes classes and two plural suffix classes.

Regular nominal class used in the model

	paradigm 1	**paradigm 2**	**paradigm 3**
Singular definite	-oo (class O)	-aa (class A)	
Plural indefinite	-ey (class E)		-awey (class L, L for long)

It means that usual base words are defined by a combination of a singular class (O or A) and a plural one (E or L). If needed, the split between singular and plural would also allow to easily integrate singularia tantum and pluralia tantum in the lexicon.

Based on this approach, there are 3426 words within the class O, 2828 in the class A, 4246 in the class E and 2062 in the class L. In case of paradigmatic variability, a base word can belong to more than two classes.

These regular models imply the removing/substitution of occurring final vowels from the base form, e.g. *gallu* > *galloo* (the final *-u* is replaced by the definite *-oo* suffix). When the combination involves an extra consonant, it is generated through other models described in 3.1.2.

3.1.2 Minor paradigms

Minor paradigms describe the sets of lemmas that require an extra consonant between the stem and the ending. Such a consonant can be a final stem consonant lost in the base form or an extra *w* consonant to avoid a hiatus after a stem in stable vowel. It concerns a limited number of base words and their compounds.

The model defines the following classes, which combine singular and plural suffixes (contrary to the ones described in 3.1.1 for which singular and plural are split):

- y: suffixes -oo/-ey with *y* as a linking consonant (2 base words).
- w: suffixes -oo/-ey with *w* as a linking consonant (6 base words).
- w: suffixes -aa/-ey with *w* as a linking consonant (8 base words).
- g: suffixes -oo/-ey with *g* as a linking consonant (for the word *doo* 'place' and it compounds only).
- g: suffixes -aa/-ey with *g* as a linking consonant (for the word *baa* 'part' only).

The classes Y, W and g involve the shortening of long final vowels before *y*, *w* or *g*.

3.1.3 Idiosyncrasies

Idiosyncrasies are directly added in the base lexicon with a shared stem field (*st:* tag) for the morphological tagger. They may involve unexpected consonant or vowel alterations. It concerns the nouns *maa* 'nom, terme', *mee* 'mouth', *moo* 'eye', *alfa* 'peasant', *ban* 'good health' as well as some pronouns.

3.1.4 Numeral specifics

Numerals follow the nominal patterns previously described, but are extended by two specific classes. The o class enables to form ordinal adjectives from (cardinal) numerals. The i class, the only prefix class, generates the numeral form prefixed by *i-*, *woy* → *iwoy* (for numbers from two to nineteen). This prefixed variation appears in standalone number forms.

3.2 The Verbal Component

In Songhay tense, mood and polarity are expressed by verbal particles, therefore verbs have no inflection. As a consequence, the Hunspell verbal component is restricted to derivational patterns.

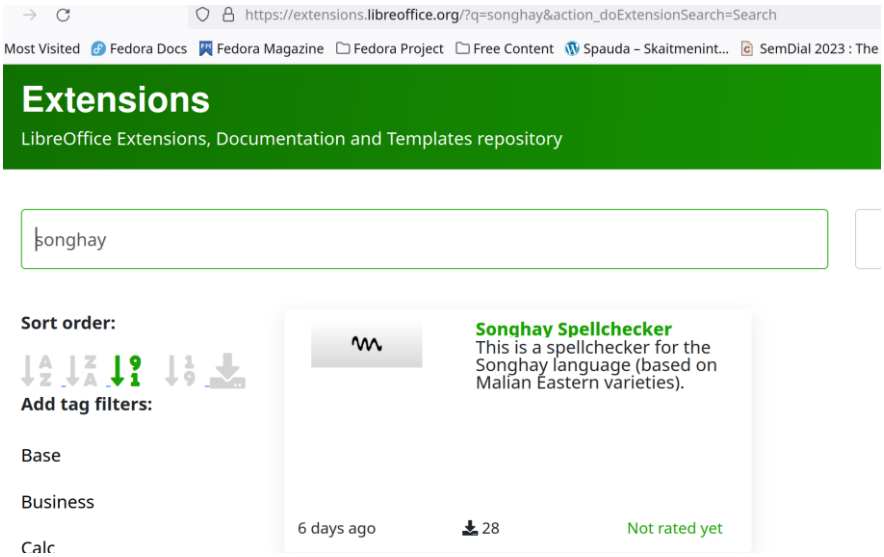
The verb affixes allow to generate the main verbo-nominal forms, the participle / verbal adjectives and the infinitive / verbal nouns. According to the nominal system, these deverbal forms are provided in singular and plural, in definite and indefinite. The verbo-nominal generation relies on the classes V and v, the latter involving a linking consonant *y* (after i) or *w* before endings, possibly with a long vowel shortening in the stem (*-uu* → *-u*, *-ii* → *-i*).

For some verbs, the generation model allows deriving the transitive/factitive suffixed counterpart in *-andi* (e.g. *naanay* 'be confident', *naanayandi* 'give confidence to sb') through classes D and d (the latter with a linking consonant as for v). This derivation pattern implies further derivation of the verbo-nominal forms for the derived transitive verb.

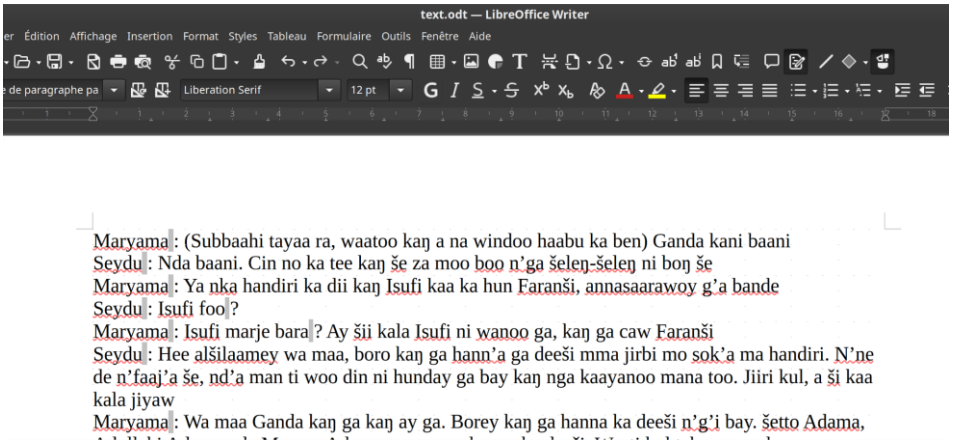
The next section illustrates the use of the Songhay Hunspell module, either packaged or not.

4. Use cases

The Songhay Hunspell Module is packed as a LibreOffice extension and can be downloaded from the dedicated webpage https://extensions.libreoffice.org/.



The spellchecker is directly usable to check the spelling of texts written using the tools provided by LibreOffice (Writer, Calc, Impress, etc.), as illustrated in the next screen capture.



The Songhay Hunspell files can also be used without the LibreOffice packaging, for browsers, for Elasticsearch or for the whole Linux system. Once installed on a Linux system, it is possible to use it for morphological analysis through the command line.

```
Maryama

Subbaahi  st:subbaahi po:noun ts:sing_indef

tayaa  st:tayo po:adj ts:sing_indef is:sing_def
tayaa  st:tayo po:noun ts:sing_indef is:sing_def
tayaa  st:taya po:adj ts:sing_indef is:sing_def
tayaa  st:taya po:noun ts:sing_indef is:sing_def

ra  st:ra po:postp

waatoo  st:waati po:noun ts:sing_indef is:sing_def

kaŋ  st:kaŋ po:pron
kaŋ  st:kaŋ po:noun ts:sing_indef
kaŋ  st:kaŋ po:verb

a  st:a po:pron is:3_sing

na  st:na po:aux
```

File sources and results files can be easily piped through Linux standard redirection operators. The output format is a verticalized text with each word form and its morphological interpretation with the lemma and the linguistic features and separated from the next word by an empty line. When the morphological analysis fails, the interpretation part remains empty (e.g. the proper name *Maryama* in the provided example). When a word form is morphologically ambiguous (e.g. *tayaa* or *kaŋ*), the word form is repeated for each possible interpretation.

5. Conclusion and Prospects

Given the lack of resources for African languages, it is important to try to find some paths to provide at least some basics for a greater number of languages with an emphasis on free tools, sharing and reusability. Hunspell module allows to fill a punctual gap for spellchecking and morphological analysis. Building on a few available language resources, namely the localization of some software interfaces and an internet lexicon on the website songhay.org, the Songhay Hunspell module is the first publicly available tool to automatically process the Songhay language. Except for some Songhay virtual keyboards, these handful of resources make up the single full digital backbone for Songhay to date. What could be the next steps?

Firstly, the Songhay Hunspell module can still be improved. The lexicon is in the process of being expanded. The analyser was run over all the texts at our disposal and the list of unknown words is currently under review. Indeed, this list contains incorrect words but also words missing in the Songhay Hunspell lexicon. After this list is thoroughly checked and purged, it is expected that a few thousands words will be added to the existing lexicon. Another expected task is to launch an initiative within the Songhay community to collect a significant list of proper nouns, mainly names, surnames and toponyms. Finally, some improvements are needed for shortened words with apostrophes, which are frequent in Songhay.

Beyond the goal of improving the Songhay Hunspell module, another step would be to provide a contextual tagger able to select one morphological interpretation for ambiguous wordforms. Together with a syntactic tagger, it would provide a very minimal set of automatic linguistic analysers, although such tools are not relevant for the everyday users of the language. But such analysers as well as other tools more directly useful for end users, like predictive typing, need at least basic task-oriented corpora. It means that in order to move further into developing tools for the Songhay community, corpus development cannot be avoided. That is why the incentive to produce written and spoken materials and to collect texts and recordings in Songhay is a fundamental task. It requires suitable tools but the challenge of mobilizing the whole language community is also of crucial importance, and it is not an easier issue than the technical one.

References:

1. Baker, M. ‘Verbal Adjectives as Adjectives without Phi-Features’. *Proceedings of the Fourth Tokyo Conference on Psycholinguistics.* Yukio Otsu (ed.), Keio University, 1-22, 2003.
2. Bailey, D. ‘Software Localization: Open Source as a Major Tool for Digital Multilingualism’. In Vannini, L., Le Crosnier, H. (eds.). NET.LANG: Toward the Multilingual Cyberspace. C&F Editions, 2012. http://net-lang.net
3. Ebongue, A. E., Hurst, E. (eds.). S*ociolinguistics in African Contexts: Perspectives and Challenges*, Switzerland: Springer International Publishing. 2017.

4. Ekkehard Wolff, H. Language ideologies and the politics of language in post-colonial Africa. *Stellenbosch Papers in Linguistics Plus*, Vol. 51, 2017, 1-22 doi: 10.5842/51-0 - 701
5. Haïdara, Y.M. *Dictionnaire soŋay-français. Kalima citaabu soŋay-annasaara senni*. Bamako: EDIS, 2010.
6. Kamwangamalu, N. M. 'Language ideologies and practices in Africa'. *Journal of Sociolinguistics*. 2019; 23: 543–554. https://doi.org/10.1111/josl.12327
7. Ògúnrẹ̀mí, T., Onyothi Nekoto, W., Samuel, S. *Decolonizing NLP for "Low-resource Languages"*, 2023.
8. Pirinen, T., Lindén, K. 'Creating and Weighting Hunspell Dictionaries as Finite-State Automata'. *Investigationes Linguisticae*. vol. XXI, 2010.
9. Osborn, D. *African Languages in a Digital Age: Challenges and Opportunities for Indigenous Language Computing*. HSRC & IDRC, 2010. Prost, A. *La langue soŋay et ses dialectes*. Dakar: IFAN, 1956.
10. Ssentanda, M., Nakayiza, J. "Without English There Is No Future": The Case of Language Attitudes and Ideologies in Uganda. In: *Sociolinguistics in African Contexts*, 2017, DOI: 10.1007/978-3-319-49611-5_7
11. Stassen, L. 'Predicative Adjectives'. In Dryer, M. S., Haspelmath, M. (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. (http://wals.info/chapter/118, accessed on 2015-05-02)