Scientific Research Publishing

Open Journal of Applied Sciences, 2026, 16(1), 1-8
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

A Semantic Wiki for Language Learning: The Case of the Baoulé Language

Lagasane Ouattara Kra[1*], Chieke Othniel Konan[1], Nouho Ouattara[1], Konan Marcellin Brou[2]

[1]Department of Mathematical Computer Science, Alassane Ouattara University, Bouaké, Côte d'Ivoire
[2]Institut National Polytechnique Félix Houphouët-Boigny, Yamoussoukro, Côte d'Ivoire
Email: *kralgasaneouat@yahoo.fr, *konanmarcellin@yahoo.fr

How to cite this paper: Kra, L.O., Konan, C.O., Ouattara, N. and Brou, K.M. (2026) A Semantic Wiki for Language Learning: The Case of the Baoulé Language. *Open Journal of Applied Sciences*, **16**, 1-8.
https://doi.org/10.4236/ojapps.2026.161001

Received: November 3, 2025
Accepted: January 1, 2026
Published: January 4, 2026

Abstract

This article presents BAOULE-WIKI, an innovative project aimed at preserving and teaching the Baoulé language through a semantic wiki that combines community collaboration and semantic web technologies. Based on the alarming findings of UNESCO and FORTIN on the imminent disappearance of African languages and the lack of suitable digital resources, the study analyses the limitations of existing initiatives (SweetWiki, Aragonais, Yoruba, Ummto) and proposes an original approach based on the automatic detection of homographs, a major problem in tonal languages. A mathematical model for recognising and eliminating homographic duplicates was designed, implemented and experimentally validated using a Python algorithm, achieving an accuracy of over 90% and a recall of around 80%, confirming the feasibility of reliable homograph differentiation. This initial contribution guarantees lexical integrity and paves the way for integration into a functional semantic wiki prototype. The results demonstrate the effectiveness of the approach while highlighting limitations related to subtle phonetic variations and multimedia integration. The article concludes with prospects for enriching the project, including the addition of audio-visual resources, extension to other Ivorian languages (Bété, Dioula, Senoufo), the use of machine learning techniques to refine tone detection, and interconnection with open knowledge bases such as Wikidata and DBpedia. Ultimately, BAOULE-WIKI aims to become a model for the digital preservation and transmission of endangered African languages. The simulation conducted on an annotated corpus yielded an accuracy of 92% and a recall of 80%, confirming the relevance of our model in the automatic differentiation of homographs. Furthermore, integrating this model into an OWL ontology and a Semantic MediaWiki prototype strengthens lexical coherence and paves the way for educational and collaborative use of BAOULE-WIKI.

Keywords

Semantic Wiki, Vocabulary, Grammar, Pronunciation, Culture, Exercises

1. Introduction

According to UNESCO, nearly half of the 6700 languages spoken around the world could disappear by the end of the century if no action is taken to promote them. In Africa, this observation is particularly alarming because, according to FORTIN [1], 95% of African languages do not have standardised digital resources (lexicons, grammars, learning tools). The Baoulé language, spoken by more than 4 million people in Côte d'Ivoire, illustrates this paradox of languages that are locally dominant but technologically invisible. Several actions have been taken to promote languages, notably by UNESCO, which designated 21 February as International Mother Language Day in 1999, and the UN, which proclaimed 2008 as the International Year of Languages. Several governments, including Côte d'Ivoire, have made the integration of languages in national education a top priority, as indicated by former Minister of Culture Maurice Bandama during the commemoration of the 3rd International Mother

Language Day on Thursday, 21 February 2013: "We must introduce national languages into education in Côte d'Ivoire. This appears to be an absolute necessity."

Faced with this challenge, collaborative technologies (wiki [2]) and the semantic web [3] offer promising opportunities for the preservation, promotion and learning of languages. It is in this context that our BAOULE-WIKI project, a semantic wiki dedicated to learning the Baoulé language, aims to combine community collaboration and semantic rigour.

2. State of the Art

Several studies have explored the use of semantic wikis, ontology, the semantic web and wikis for knowledge capitalisation and language learning. Among the most notable projects are:

- **SweetWiki**: Improving organisation, navigation and search within wikis by leveraging semantic web technologies, including a user-friendly interface (WYSIWYG editor), semantic annotation via RDFa (Resource Description Framework in attributes) and the use of folksonomies [4].
- **Aragonais Project**: Documenting a Spanish minority language via Semantic MediaWiki (SMW) by enabling active community collaboration and simple structuring [5].
- **Wiki Ummto**: Wiki Ummto is a semantic wiki that enables the capitalisation of disciplinary knowledge for teaching using semantic web technologies, Servlets and JSP [6].
- **Onto Collab** [7] (2015): Co-construction of ontologies via a distributed architecture, adapted to collaborative projects by facilitating joint work between

business experts and knowledge engineers, but also by integrating a strategic review system.

- **Yoruba Project**: Modelling the tones of an African language using RDF properties, with explicit consideration of tones in the ontology and the possibility of SPARQL queries targeting phonological characteristics [8].

Baoulé has complex linguistic characteristics, including a five-level tonal system (high, low, mid, rising, falling) that directly affects the meaning of words. The lack of standardized spelling and the presence of dialectal variations add difficulties to the numerical structuring of the language. These specificities explain the frequent occurrence of homographs, the automatic detection of which is an essential prerequisite for any reliable semantic modeling.

However, these approaches have common limitations:

- Lack of management of dialectal variations and semantic ambiguities.
- Lack of adaptation to tonal languages.
- Complex and inaccessible interfaces.
- Poor multimedia integration (audio, video).

3. Proposed Approach

The aim of our contributions is to set up a semantic wiki capable of automatically distinguishing between homographs, in order to eliminate duplicates linked to identical spelling but different pronunciation and meaning.

Our approach is based on two major contributions:

1) Design and simulation of a mathematical model for recognising and eliminating homographic duplicates based on linguistic analysis.

2) Implementation of a prototype web application for a semantic wiki integrating the mathematical model.

The choice of Semantic MediaWiki (SMW) is based on its ability to combine an accessible collaborative interface with a powerful semantic core based on RDF and OWL. SMW also allows the integration of structured forms, the automatic generation of RDF triples, and the querying of data via SPARQL, making it particularly suitable for a multilingual system requiring semantic validation of entries.

4. Proposed Resolution

4.1. Mathematical Model for Homograph Detection

The objective of this contribution is to develop a mathematical model for recognising and eliminating homographic duplicates based on linguistic analysis.

Baoulé and French, like many tonal languages, contain numerous homographs: words with identical spelling but different meanings and pronunciations (e.g. KO = “to go” or “lord”, avocat = “fruit” or “profession”). To solve this problem, we have designed a formal mathematical model that automatically detects and differentiates between these homographs in order to eliminate homographic duplicates and ambiguities.

Let W be a set of words such that for each word $w_i \in W$,

$$S(w_i) = \{s_1, s_2, \cdots, s_n\}.$$

where: $fonction\ d'association\ de\ type\ \ S : W \to P$ (sens) (it associates each word $w_i \in W$ to a subset of possible meanings). Formally, we have:

$$\begin{aligned} \text{graph}(w_i) = \text{graphie}(w_j) \wedge \left[S(w_i) \neq S(w_j) \right] \vee \text{categorie}(w_i) \\ \neq \text{categorie}(w_i) \vee \text{prononciation}(w_j) \\ \neq \text{prononciation}(w_j) \Rightarrow \text{homographes} \end{aligned}$$

We can represent each word w by a characteristic vector V with four parameters representing the criteria for distinguishing a word w:

$$V(w) = (\text{graphie, prononciation, catégorie, sens})$$

In our implementation, pronunciation is encoded in a categorical form according to five standardized tonal levels (high, low, middle, high-low, low-high). Meaning is represented as a unique identifier associated with a validated lexical definition. The similarity score used in the results corresponds to a normalized comparison between the components of the vector $V(w)$, allowing for the measurement of lexical proximity between two entries.

Consequently, we can define a detection function H:

$$H(w_1, w_2) = \begin{cases} -1 & \text{si graph}(w_1) \neq \text{graphie}(w_2) \\ 1 & \text{si graphie}(w_1) = \text{graphie}(w_2) \wedge V(w_1) \neq V(w_2) \\ 0 & \text{si non} \end{cases}$$

In our case:

- If $H(w_1, w_2) = -1$ → it is a new word, so record it.
- If $H(w_1, w_2) = 1$ → the two words are homographs, so they are recorded with a unique URI for each word, allowing them to be distinguished.
- If $H(w_1, w_2) = 0$ → these are exact duplicates, so the record is cancelled.

A comparison algorithm was implemented in Python to identify homographs and generate unique URIs (e.g., baoule:kɔ:verb:to go) and also to empirically validate our mathematical model for detecting homographs and eliminating duplicates.

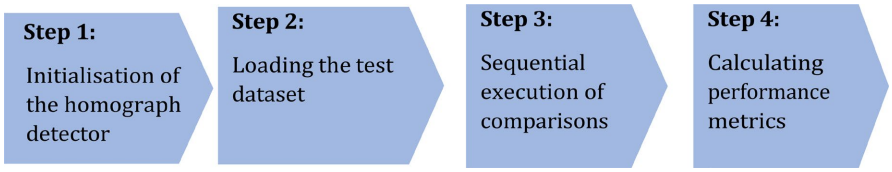
4.2. Ontological Modeling

The BAOULE-WIKI ontology, designed in OWL2, formalizes essential linguistic concepts: classes (Word, Tone, Meaning, Grammatical Category, Dialect), properties (hasTone, hasCategory, hasMeaning, isHomographOf), and integrity rules. This ontology serves as a semantic infrastructure to ensure data consistency, facilitate the automatic detection of homographs, and enable advanced queries for learning purposes

5. Results and Discussions

We simulated our mathematical model in order to evaluate its performance in detecting and recognising homographs. The simulation process consisted of four

steps:



For our test dataset, we collected a set of Baoulé words from a native speaker, combined with certain French words, in order to test our mathematical model and demonstrate the multilingual aspect of our system.

A detailed comparison of homograph detection is summarized in Table 1 below.

Table 1. Detailed results of homograph detection.

No.	Word 1	Word 2	Expected result	Result obtained	Similarity score
1	kô (verb)	kô (noun)	Homograph	1	0.0
2	Lawyer (noun)	Lawyer (noun)	Homograph	1	0.0
3	Convent (verb)	Convent (noun)	Homograph	1	0.0
4	Li (noun)	Li (adjective)	Homograph	1	0.0
5	sè (verb)	sè (verb)	Duplicate	0	1.0
6	bla (adjective)	bla (adj)	Duplicate	1	0.0
7	kô (verb)	yê (interjection)	Different	−1	-
8	kô (noun)	Wè (conj)	Different	−1	-

The expected results in Table 1 were obtained except in one case, pairing 6, where the result obtained did not match the desired result. This error can be explained by an initially incomplete annotation of the dataset: the two entries for the word bla shared the same spelling but lacked sufficiently explicit semantic or tonal distinctions, leading the algorithm to consider them as homographs rather than duplicates. This highlights the need for phonetic enrichment and rigorous normalization of annotations.

Indicates that when the algorithm predicts a homograph, it is rarely wrong; the slightly lower recall reflects a few borderline cases where subtle differences (e.g., tone vs. minor spelling variation) may require normalisation or enrichment (e.g., more explicit phonemic properties). In conclusion, we can say that these results demonstrate the feasibility of automatic homograph differentiation in a tonal language context.

As illustrated in Figure 1, the detection model achieved a performance accuracy of 92%, with a recall rate of 80%.

- High accuracy (>90%): the algorithm correctly identifies homographs without false positives.

- Slightly lower recall (~80% - 85%): a few borderline cases occur when the tone difference is subtle. In other words, high accuracy on "Homograph".

The classification performance metrics are presented in Table 2.

Table 2. Classification metrics (Homograph vs Non-homograph).

Class	Accuracy	Recall	F1 score	Support
Non-homograph	75	94%	86%	3
Homograph	92	80	89	5
Accuracy			88	**8**

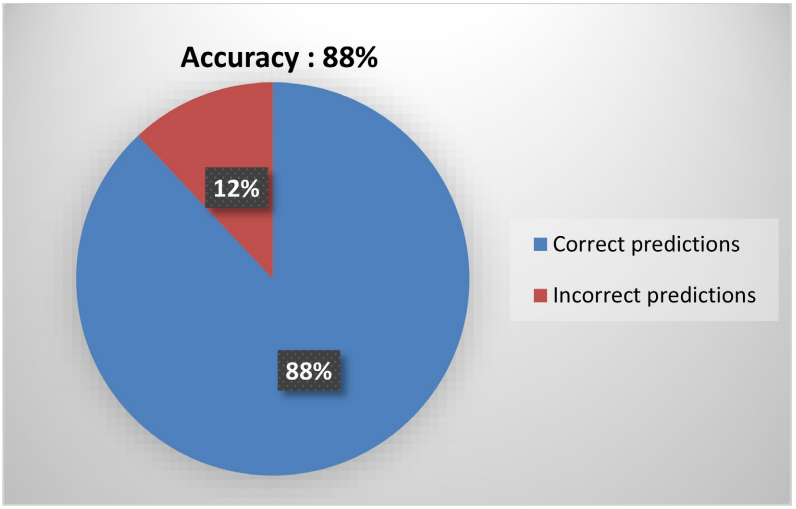


Figure 1. Diagrams illustrating the performance of the detection model.

The results obtained confirm the relevance of our mathematical model for solving the central problem of homographs for tonal languages. Although the accuracy rate of 92% shows that our model is highly reliable in flagging many homographs and avoiding many false positives, it would guarantee data integrity in a collaborative system. This result indicates that our work demonstrates significant progress compared to what has already been done, for example, Yoruba, which had a consistent degree of homogeneity (tone problem) but did not propose any automated sub-mechanisms. It should be noted, however, that the 80% recall rate shows that our model sometimes fails to flag certain homographs. For example, when the tonal and semantic variations of sounds are subtle. This limitation, often linked to the fineness of manual annotations or dialectal variations not yet modelled, highlights the need to enrich the model with more advanced layers of phonetic analysis, potentially via machine learning. Thus, while our approach goes beyond the limitations of traditional semantic wikis by offering an operational solution for a tonal language, it also paves the way for future improvements to capture the full complexity and richness of the Baoulé language.

These results demonstrate the relevance of our model but also reveal that accuracy depends heavily on the quality of the lexical annotation. Future improve-

ments will involve adding more detailed phonemic attributes, integrating audio files for tones, and combining the current model with supervised learning techniques.

6. Conclusion and Outlook

BAOULE-WIKI offers an innovative solution for the preservation and learning of Baoulé, combining community collaboration and semantic technologies. The homograph detection model empirically validates our approach and paves the way for reliable capitalization of linguistic knowledge.

The next phase of the project includes the integration of multimedia elements (audio, video), the expansion of the ontology to Baoulé dialectal variants, and the use of machine learning algorithms to improve the detection of complex tones. Finally, connecting to open knowledge bases such as Wikidata and Lexvo will allow BAOULE-WIKI to be incorporated into the semantic web ecosystem.

Outlook:

In the short term, we plan to:

- Completing the development and deployment of BAOULE-WIKI for simulation and evaluation by all users.
- Enriching multimedia resources (audio recordings, videos).
- Extending the model to other Ivorian languages (Bété, Senoufo, Dioula).
- Integrating machine learning techniques to improve the detection of phonetic variations.
- Connecting to open linguistic databases (Wikidata, DBpedia).
- Educational integration: implementation of personalized learning paths, interactive exercises and automated assessment of learning outcomes.
- Community participation: online deployment with a community validation module to enhance data reliability.

In the long term, BAOULE-WIKI could serve as a model for the preservation of other endangered languages, thereby contributing to global linguistic diversity.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Fortin, A. (2022) UNESCO's Digital Education Projects in French-Speaking Africa: Objectives, Challenges and Prospects. https://archipel.uqam.ca/16209/1/T1116.pdf

[2] Ebersbach, A., Glaser, M., Heigl, R. and Warta, A. (2008) Wiki: Web Collaboration. Springer Science & Business Media.

[3] Buffa, M., Ereteo, G. and Gandon, F. (2007) Wiki and Semantic Web. https://www.researchgate.net/publication/29606923_Wiki_et_Web_Semantique

[4] Buffa, M., Gandon, F., Ereteo, G., Sander, P. and Faron, C. (2008) SweetWiki: A Semantic Wiki. *Journal of Web Semantics*, **6**, 84-97. https://doi.org/10.1016/j.websem.2007.11.003

[5] Garabato, C.A., Boyer, H. and Calvet, C. (2023) Languages on the Verge of Substitu-

tion and the Glottotherapies Applied to Them (Aragonese, Occitan).

[6] Sara, Z. and Zina, T. (2016) Use of a Semantic Wiki for Teaching Management. Mouloud Mammeri University.

[7] Pushpa, C.N., Deepak, G., Thriveni, J. and Venugopal, K.R. (2015) Onto Collab: Strategic Review Oriented Collaborative Knowledge Modeling Using Ontologies. 2015 *Seventh International Conference on Advanced Computing* (*ICoAC*), Chennai, 15-17 December 2015, 1-7. https://doi.org/10.1109/icoac.2015.7562785

[8] Okediya, T., Afolabi, I., Iheanetu, O. and Ojo, S. (2019) Building Ontology for Yorùbá Language. *Proceedings of the First International Workshop on NLP Solutions for Under Resourced Languages* (*NSURL 2019*), 124-130.