International Journal of Computer Science and Mathematical Theory (IJCSMT) E-ISSN 2545-5699
P-ISSN 2695-1924 Vol 11. No. 6 2025 www.iiardjournals.org

Development of Digital Lexicon for Etsako Language.

Osumah Thomas O.
Department of Computer Science,
Federal College of Education, Okene
Kogi State, Nigeria.
Osigbe16@gmail.com
DOI: 10.56201/ijcsmt.vol.11.no6.2025.pg1.6

Abstract
The preservation of indigenous languages has become increasingly important in today digital world. Etsako language is a low resource language spoken by the people of Edo North in Nigeria. The development of this language digitally is posed with challenges such as lack of standard orthography, bilingual lexicon for performing natural language processing tasks, declining usage by Etsako indigenes and limited digital representation on the web. This paper explores the development of digital bilingual lexicon required for performing machine translation tasks, language documentation and e-learning tool for teaching and learning Etsako language. The research highlights the method used, which include the curation of over 5000 lexicon datasets developed digitally and stored on a cloud storage. A web-based digital lexicon application for Etsako language was developed. Bootstrap 5 and PHP and MySQL were adopted as software development tools. The developed lexicon will enable users to generate English text and its equivalent translation in Etsako language, while the dataset will enable researchers to develop natural language processing applications for Etsako language.

Keywords: *Bilingual Lexicon, Machine Translation, Etsako Language, Parallel Corpus, Natural Language Processing (NLP).*

1. Introduction

Developing language resources like lexicons and parallel corpora is crucial for various natural language processing tasks including machine translation, sentiment analysis, information extraction, text generation, text summarization, and question answering. Natural Language Processing (NLP) focuses on using computing devices to comprehend and produce human languages (Chinenyeze, et.al. 2019). Researchers in Computer Science, Artificial Intelligence (AI), and Computational Linguistics should prioritize creating digital bilingual lexicons and annotated parallel corpus for every language globally to prevent their extinction. While some languages have been well-represented on the web, several minority languages face extinction due to their lack of digital linguistic resources that are not available on the web (Imelda, 2021).

Nigeria, with over 500 local languages (Matt, 2018), has only three national languages (Igbo, Hausa and Yoruba) fully represented on Google Translate. This discrepancy creates a gap between high-resource and low-resource languages like Etsako language. To address this, there is a critical need to enhance electronic resources for Nigerian languages (Olutola, et.al. 2015), specifically through web 2.0 tools for minority languages by creating standard orthography, bilingual lexicon and parallel corpus for each of the local languages.

The recent approval of a new national language policy in Nigeria, emphasizing mother tongue in education, faces challenges of proper implementation in basic education due to the limited representation of local languages online (Deji, 2022). Etsako language, spoken by people in three local government areas in Edo State, lacks sufficient learning resources, posing obstacles to its use and preservation. Implementing web 2.0 technologies can bridge this gap by enabling the creation and sharing of Etsako language content for learners and speakers. By utilizing platforms like Facebook, WhatsApp and YouTube, developers and learners can connect and access multimedia educational content in Etsako language.

2. Preservation of Etsako Language in Digital Form

Every community tries to retain their cultural values by teaching and sharing it from one generation to another through the language used as a means of communication. Onyemelukwe (2019) described language as the vehicle used for the transmission of the culture of an ethnic group from one generation to the other, which lives and its continuous use by the human group serves as a means of sustainable growth.

Studies carried out on languages (Eludiora, et.al. 2016; Etuk et al., 2020; Chinenyeze, et.al. 2019) considered as endangered and threaten for extinction show that their identity, instrument of communication, linguistic values, historical activities and social reconstruction are bound to be lost in the nearest future if nothing is done to prevent them from happening.

Etsako language is one of the minority languages in Nigeria currently considered as endangered and threaten for extinction (Imelda, 2021; Adeoye, 2012), as the current speakers of the language are mainly adults that are residing in the villages. This language also lacks standard orthography and digital lexicon needed to develop the language digitally as well as a tool for other natural language processing tasks. Imelda (2021) noted that Nigerian languages should have formal spelling systems that will guide their development.

Hausa, Yoruba, Igbo and other foreign languages have been integrated into Google Translate Engine to help individuals and organizations overcome language barrier through a machine translation technique. Oge (2022) reported that Google Translate adds 24 new languages but no Nigerian language is included this time in a list that already has Igbo, Yoruba and Hausa.

Therefore, the need to develop digital resources for Etsako language that will address the lack of standard orthography, represent the Etsako language in the public domain and preserve it from total extinction via web 2.0 tools such as WhatsApp platform. This paper explores the effort made to develop digital resources such as bilingual lexicon and machine translation application to provide easy access for Etsako language.

3. Literature Review

In Nigeria, English language holds significant sway in business dealings both nationally and globally, as well as in religious institutions, schools and offices. Omoregbe, et.al. (2014) noted that there is a competitive dynamic between English and indigenous Nigerian languages, with the latter striving for survival. This rivalry is exacerbated by the widespread use of mobile technology, predominantly featuring English content. Mobile technology has revolutionized individual and corporate activities through digitalization. Numerous studies have explored foreign languages and key Nigerian languages like Hausa, Yoruba, and Igbo (Eludiora, et.al. 2015), contributing to the enhancement and preservation of linguistic resources crucial for manipulating these languages. However, low resource languages such as Etsako face extinction due to a lack of standardized

orthography hindering educational and research advancements, jeopardizing the rich cultural heritage and history embedded in these languages.

Every community's language serves as a vessel for its traditions, past, and values, emphasizing the importance of preserving these linguistic identities, as highlighted by Imelda & Ogbonna (2015). Their insights underscore the intrinsic value of each language irrespective of speaker population, emphasizing the role languages play in preserving cultural legacies.

The web acts as a vast repository of information for both learners and educators, offering a platform for accessing diverse educational materials. Web technologies have evolved from the read-only Web 1.0 to the interactive read-and-write Web 2.0, transforming the web into an engaging learning space. Web 2.0 technologies, encompassing platforms like WhatsApp, Wikipedia, YouTube, Instagram, and Facebook, have significantly enhanced knowledge acquisition and academic performance. These tools have globalized learning environments, providing interactive and dynamic platforms for educational engagement. Web 2.0 tools empower users to craft personalized web pages, share multimedia content, and engage in collaborative learning. Social media platforms integrated with Web 2.0 enhanced educational interactions among teachers, students and parents to facilitate seamless information exchange. In this context, WhatsApp has been used to develop Etsako language, leveraging virtual forums for content creation and validation and utilizing audio-visual resources for language learning and preservation.

4. Material and Method

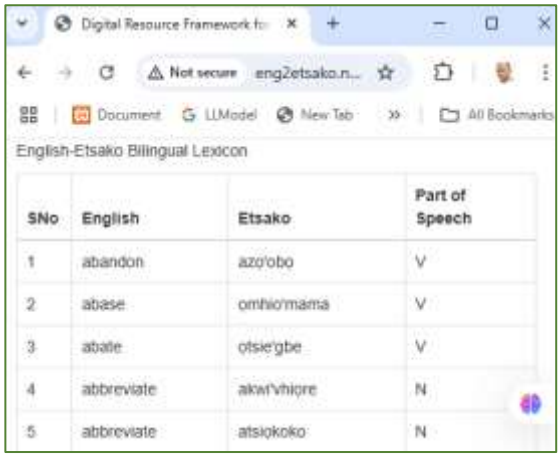
In developing digital lexicon for Etsako language, current hard copy of Etsako language text were reviewed (Adamu & Audu, 2011; Osumah, 2013; Enaboakpe, 2016; Etu, 2019). Etsako language speakers and experts were consulted to help curate the dataset for building the digital lexicon. WhatsApp as a popular web 2.0 tool was adopted to create a virtual forum in order to reach out to wider audience for curating, translating and validation the various dataset used in the development of the digital lexicon for Etsako language. GitHub cloud storage was used to host the dataset of the digital lexicon for Etsako language.

The software application was developed using Bootstrap 5.0 for the graphic user interface, while PHP was used as the programming tool and MySQL as the database tool. A web-based digital lexicon for Etsako language has been developed.

5. Result

The digital bilingual lexicon for Etsako language has been developed as a web-based application and linguistic resource created and stored in digital format needed for natural language processing tasks and other research work in computational linguistic. The develop resources consist of dataset formatted as comma separated value file stored and hosted in GitHub cloud storage and web-based application hosted on the web (www.eng2etsako.ng).

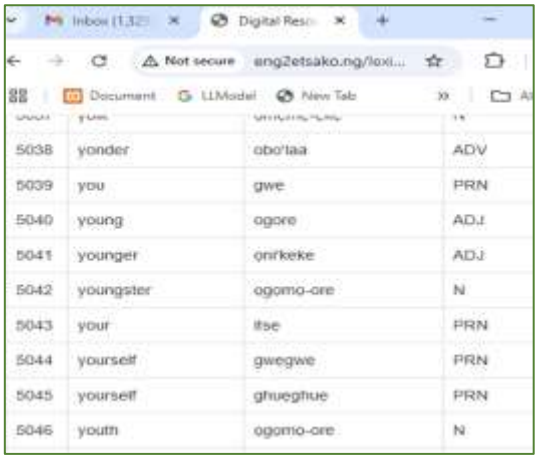
The English - Etsako Bilingual Lexicon application, features over 5,000 English words and Etsako translation equivalent alongside with part of speech (POS) arranged alphabetically, as represented in figure 1 and figure 2 below. Figure 1 displays the first set of five records while figure 2 shows the last set of records in the lexicon database.



English-Etsako Bilingual Lexicon

SNo	English	Etsako	Part of Speech
1	abandon	azo'obo	V
2	abase	omhio'mama	V
3	abate	otsie'gbe	V
4	abbreviate	akwi'vhiore	N
5	abbreviate	atsiokoko	N

Figure 1 – *First set of records in Etsako Lexicon*



5038	yonder	obo'taa	ADV
5039	you	gwe	PRN
5040	young	ogore	ADJ
5041	younger	onirkeke	ADJ
5042	youngster	ogomo-ore	N
5043	your	itse	PRN
5044	yourself	gwegwe	PRN
5045	yourself	ghueghue	PRN
5046	youth	ogomo-ore	N

Figure 2 – *Last set of records in Etsako Lexicon*

The application also has a module - *Data Entry and Word Search*. The data entry form module is the interface where the user can create new record and append into the lexicon database. The word search enables the user to search for an existing lexical item from the database.

6. Discussion

The absence of online lexicon for Etsako language on the web has led to the development of digital bilingual lexicon for performing natural language processing tasks and language documentation for Etsako language. This effort shows a significant step towards the preservation and revitalization of Etsako language as one of the low resource languages in Nigeria.

The curation and storage of over 5000 bilingual lexicon datasets (stored as CSV file) of English text, part of speech and their equivalent translation in Etsako language that are currently hosted on the web as well as on Github cloud storage will enable users to generate English text and Etsako translation via Artificial Intelligences (AI) tools.

The developed digital lexicon has provision for appending new data items that are not currently available in lexicon dictionary, as well as the ability to edit existing data items either deleting or correcting wrong spelling. The lexicon will serve future researchers in performing natural language processing tasks and a tool for documenting Etsako language. It will also serve as an e-learning tool to learn, read and write Etsako language.

7. Challenges of Bilingual Lexicon Development for Low Resource Languages

In developing bilingual lexicon for Etsako language as one of the low resource languages in Nigeria, several challenges were faced in the digitization process. Some of these challenges include:

Lack of Standardized Orthography: There are currently no standard orthography available for Etsako language as the text material consulted have variations in spelling and pronunciation. Bird (2020) noted that some languages are primarily spoken and lack standardized written forms, making it challenging to create linguistic resources.

Limited Data Availability: There are currently insufficient written texts prepared in digital format that can be used to perform natural language processing tasks. In their view, Joshi et al., (2020) stated that many low-resource languages lack the large-scale annotated corpora, dictionaries,

parallel texts, and linguistic databases that modern natural language processing (NLP) systems depend on.

Inadequate linguistic experts: There are insufficient linguists for Etsako language to help develop its linguistic resources and store in digital format for natural language processing tasks.

Technological Barriers: The high cost and maintenance of computing devices as well as unstable power supply created barrier for access to technology among native speakers of Etsako language to help develop the language digitally (Akande, 2020). Ogunbiyi (2018) also stated that many community centers lack computing devices or reliable internet access, thereby affecting language learners' opportunities to engage with digital resources.

Lack of resource development centre in tertiary institutions: There are currently no digital resources development centres for most minor or low resource languages in Nigeria. For example, Etsako language has no linguistic development centre in any of the tertiary institutions including Federal Polytechnic, Auchi; Ambrose Alli University, Ekpoma and Edo University, Iyamo where the language is domiciled.

8. Conclusion

The need to develop digital resources such as bilingual lexicon and annotated corpus for Etsako language is important, as this effort will lead to the preservation and its proper documentation of this language for its users from time to time. This study explained the approach taken to develop digital bilingual lexicon creation and web-based lexicon application for Etsako language, thereby addressing linguistic and technological challenges affecting the language development. Etsako language in terms of digital resources (needed for performing Natural Language Processing (NLP) tasks) have not been fully represented on the web before now, as this study will ensure the successful integration of Etsako language into the digital landscape.

It is imperative that there should be collaboration between linguists, information technology researchers and native speakers of Etsako language to sustain its development digitally like other foreign and national languages in Nigeria. Tertiary institutions where Etsako language is domiciled should setup “Etsako Language Resource Development Centre” to aid researchers in computing disciplines and computational linguistics who may wish to carry out studies on NLP tasks for Etsako language.

9. Future Direction of the Research

More efforts will be put in place to update and expand the current bilingual lexicon to accommodate voice (speech) data items, sentence examples for each record. Machine translation system will be developed for Etsako language to aid the speakers of the language learn and utilize the language. A standard parallel corpus for Etsako language will also be developed to accommodate the use of large language model and neural machine translation approach to develop a more robust translator for Etsako language. This will also help to address the challenge of word sense disambiguation for the language.

References

Adeoye, A. (2012). Web-based English to Yoruba noun-phrases machine translation system Master's Thesis, Federal University of Technology (FUTA), Akure.

Akande, A. (2020). Challenges of ICT Integration in Nigerian Schools: Lessons from Etsako. *Journal of Digital Education*, 5(3), 45-59.

Bird, S. (2020). Decolonising Speech and Language Technology. Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020).

Chinenyeze C. E, Bennett E. O. and. Taylor O. E. (2019). A Natural Language Processing System for English to Igbo Language Translation in Android, *International Journal of Computer Science and Mathematical Theory*, Vol. 5(1).

Eludiora, S.I, Afolabi B.O and Sanusi, A.A (2016). Development of english to hausa machine translation system, *Research Journal of Engineering and Technology* 2 (1), pp. 12-32

Eludiora, S.I, Agbeyangi, A.O and Fatusin O.I (2015). Development of English to Yorùbá Machine Translation System for Yorùbá Verbs' Tone Changing *International Journal of Computer Applications.*

Eludiora, S.I and Odejobi, O.A (2016). Development of an english to yorùbá machine translator, *International Journal of Modern Education and Computer Science (IJMECS),* 8(11).

Etuk S. O., Oyefolahan, I. O., Zubairu, H. A., Babakano, F. J., and Momoh, L. O. (2020). Development of an English-to-Ebira and Ebira-to-English Machine Translation System**,** *International Journal of Information Processing and Communication (IJIPC)* Vol. 10 (1 & 2).

Enaikele, M.D., (2014). Etsako: an anthropological reflection of an endangered minority language in Nigeria. *Journal of Pan African Studies*, Vol. 7(4).

Imelda, U. (2021). The Empowerment of Nigerian Languages for Sustainable Development. https://www.researchgate.net/publication/368288988.

Imelda U., & Ogbonna A.,(2015). Protection and promotion of indigenous Nigerian languages, Awka *Journal of Linguistics and Languages (AJILL)* Vol. 9(1).

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The State and Fate of Linguistic Diversity and Inclusion in the NLP World. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics

Matt, S. (2018). List of Languages in Nigeria, Available at www.naijahomebased.com

Oge, U. (2022. Google Translate adds 24 new languages. Premium times, May 12, 2022. https://www.premiumtimesng.com/business/business-news/529241-google-translate-adds-24-new-languages.html.

Ogunbiyi, O. (2018). The Digital Divide and Language Development in Nigeria. Linguistic Research, 32(2), 123-137.

Olutola, F., Akinwale O., Kayode A. & Boniface A (2015). Digital Yorùbá Corpus *International Journal of Innovative Science, Engineering & Technology*, 2(8). Available at: www.ijiset.com

Omoregbe, N.A, Ambrose A. A, Adewole A & Olayemi O. O. (2014). Design and implementation of Yoruba language mobile tutor. Proceedings of EDULEARN14 Conference, 7th-9th July, Barcelona, Spain.

Onyemelukwe, I.M. (2019). Language endangerment: The case of the Igbo language. *IMPACT: International Journal of Research in Humanities, Arts and Literature* (IMPACT: IJRHAL), ISSN (P): 2347-4564; ISSN (E): 2321-8878, 7(1).