

Kennesaw State University

## DigitalCommons@Kennesaw State University

---

African Conference on Information Systems  
and Technology

The 10th Annual ACIST Proceedings (2024)

---

Sep 12th, 11:00 AM - 11:25 AM

### Duramazwi Digital System (Modules Integration)

Laiza Mumiriki

*Chinhoyi University of Technology*, mumirikilaiza@gmail.com

Surprise Prosper Chikotsi

*Chinhoyi University of Technology*, prosperchikotsi@gmail.com

Tafadzwa Muchaiwa

*Chinhoyi University of Technology*, tmuchabaiwa@gmail.com

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/acist>



Part of the [African Languages and Societies Commons](#), and the [Technology and Innovation Commons](#)

---

Mumiriki, Laiza; Chikotsi, Surprise Prosper; and Muchaiwa, Tafadzwa, "Duramazwi Digital System (Modules Integration)" (2024). *African Conference on Information Systems and Technology*. 1. <https://digitalcommons.kennesaw.edu/acist/2024/presentations/1>

This Event is brought to you for free and open access by the Conferences, Workshops, and Lectures at DigitalCommons@Kennesaw State University. It has been accepted for inclusion in African Conference on Information Systems and Technology by an authorized administrator of DigitalCommons@Kennesaw State University. For more information, please contact [digitalcommons@kennesaw.edu](mailto:digitalcommons@kennesaw.edu).



# Duramazwi Digital System (Modules Integration)

## Research Paper

**Tafadzwa Muchabaiwa**  
Chinhoyi University of Technology  
[tmuchabaiwa@cut.ac.zw](mailto:tmuchabaiwa@cut.ac.zw)

**Laiza Mumiriki**  
Chinhoyi University of Technology  
[mumirikilaiza@gmail.com](mailto:mumirikilaiza@gmail.com)

**Surprise P Chikotsi**  
Chinhoyi University of Technology  
[prosperchikotsi@gmail.com](mailto:prosperchikotsi@gmail.com)

## ABSTRACT

This research paper explores the development of the Duramazwi Digital System, a comprehensive platform that integrates digital tools and technologies to enhance the experience of linguists and language learners. By combining the use of a dashboard, API, mobile app, and chatbot to provide users with seamless and interactive language learning experience. The study adopted the Activity Theory as its guiding theory for theoretical insights on natural language processing systems, language and learning. The methodology employed in this research follows the Design Science Research approach. The study utilizes various data collection techniques. The paper presents the system design and implementation details, including identification of milestones, requirement analysis, and development of various system diagrams. The study concludes by highlighting the strengths and benefits of the Duramazwi Digital System, such as improved performance, availability, maintainability, usability, user-friendliness, and security. The paper also acknowledges the weaknesses and provides recommendations for future maintenance and improvements.

## Keywords

Dictionary, Duramazwi, Lexicography, Multilingual societies, Shona, Translation

## INTRODUCTION

The term Duramazwi (*dura* means to review, *mazwi* means words) (Chimhundu H. , 2005) is a word rendered from English word dictionary. It is a dictionary of Shona in the Shona language. The structure

of the Shona language encompasses noun Classes, verb structure, tonal system, noun and verb concordance and word order, all which make it difficult to simply adopt Natural Language Processing (NLP) algorithms that have been used for the English language dictionaries and form the bedrock of Large Language Models. It emphasizes the importance of including both traditional and contemporary vocabulary, as well as providing detailed explanations of word usage and cultural contexts (Emmanuel, 2006). These cultural characteristics include sensitive and offensive words, words with long meanings, relational norms, cultural norms, proverbs and idioms. These reflect the Shona way of life, beliefs, customs and traditions. Only verbal aspects can be captured in a dictionary, but in Shona there are many non-verbal cultural aspects in everyday life (Chimhundu & Chabata, 2007). Furthermore, this highlights the standard orthography of Shona, making it easy to establish consistent spelling and pronunciation rules. The existing Shona dictionaries do not have speech recognition, translation, summarisation and paraphrasing at the same system. To solve this problem, the Duramazwi Digital System was divided into four different modules, which are module 1 -speech recognition, module 2 -search, question answering and dictionary, module 3 -translation and summarisation and module 4 - modules integration. Module 2 - search, question answering and dictionary, being the main module that focuses on the creation a database that is used by modules 1- speech recognition and 3 - translation and summarisation. The study adopted the Activity Theory rooted in the work of Lev Vygotsky and further developed by scholars like Engeström, which is a framework that examines human practices as developmental processes, with both individual and social levels interlinked.

## **BACKGROUND**

The Shona language was one of the major Bantu languages spoken by the Shona people in Zimbabwe. The Shona language has a history of about a century and a half. This goes back to the time before the arrival of the Pioneer Column and the establishment of Southern Rhodesia by the South African Company, when Shona language texts were found in English and in collections. Shona words, as the first bilingual dictionary in Shona and English. Since 1959, it has been a trend to create bilingual dictionaries. The most famous of these dictionaries is Hannan's Standard Shona Dictionary, published in 1959 and revised in 1984. Bilingual dictionary compilation continued until 1996, when it was created by the African Languages Lexical Project (ALLEX), housed in the Department of Languages and Literatures and Africa at the University of Zimbabwe, published the first monolingual Shona dictionary, Duramazwi re ChiShona. This was followed in 2001 by Duramazwi Guru re ChiShona, a monolingual Shona dictionary. Since the publication of this dictionary, the nature of Shona dictionaries has changed, from creating general language dictionaries to compiling specialized dictionaries. Researchers such as Bisandu have explored the need for Duramazwi Digital that covers different languages and registers. They emphasized the importance of incorporating traditional and contemporary vocabulary, and provide explanations of vocabulary usage and cultural contexts.

## **PROBLEM STATEMENT**

The importance of incorporating traditional and contemporary vocabulary, and provide explanations of vocabulary usage and cultural contexts. Previous studies have shown that there was no written record information Shona, making it difficult to establish spelling and pronunciation rules, the lack of

Integrated modules between speech recognition, translation, summarisation and paraphrasing was hindering the seamless flow of information processing and limiting user experience.

## **LITERATURE REVIEW**

This section looks at various studies and previous work on important concepts that could be used when carrying out projects. Although the literature covers a broad range of these theories, this analysis will focus on four crucial applications that are frequently discussed on a global level in the literature under consideration.

### **Evolution of Dictionaries**

Over time, electronic dictionaries have evolved from standalone units to software and mobile applications that provide more content at lower cost. The Electronic dictionary was first introduced in the early 1970s. The main users of these versatile and expensive tools are students and language experts. Previously these were standalone devices with disk drives or word-based cartridges. In 1975, one of the first electronic dictionaries called Lexicon 2001 was published. It's a big, heavy piece of equipment that would cost \$1,000 in today's money. Lexicon 2001 allowed users to search for words using a keyboard and a cathode ray display (Kerssens, 2016). With the advent of personal computers in the 1980s, electronic dictionaries began to transition from standalone devices to software applications. Many dictionary publishers started creating electronic versions of their dictionaries, which could be installed on personal computers (Rundell & Stock, 1992). They also began to offer additional features, such as thesaurus functionality, pronunciation guides, and word games. Mobile apps also offer additional features, such as speech-to-text translation, voice-enabled search, and offline access (Benbada & Benaouda, 2023).

### **Dashboard**

This design has several stages including the program structure, admin dashboard design and the main page in development (Susilana, Gema, Ardiansah, & Yuyu, 2022) concept content in education and learning sourced from primary references was considered to be limited whereas our web-based interface that provides an overview of the dictionary platform's key performance indicators, user statistics, and other relevant data. It allows administrators and managers to visualize and analyse data, monitor user activity, and make informed decisions based on the insights gained from the data. The dashboard can display information such as user engagement metrics, search trends, popular words or phrases, and user feedback.

### **Application Programming Interface**

The API acts as the bridge between the dictionary platform and external applications or services. Developers can use the API to access linguistic data from the platform and incorporate it into their own products. For example, a language learning app could utilize the API to fetch definitions and examples from the dictionary platform to enrich its educational content. The API structure allows the user to get a lot of information from one place, rather than relying on different tasks and manual processing of existing information to get more information for the keyword. Instead of receiving a large list of

individual values, users select a word they want to access and receive all the grammatical, syntactic, and semantic information associated with that word. In this context, we value quality rather than quantity (Lonke, Ilan, & Dzhuranyuk, 2022).

### **Mobile Application**

Mobile dictionary applications are widely available and used for mobile phones. Students can download various dictionaries with audio functions to their phones to master new expressions. Portable dictionaries Oxford Dictionary, Merriam Webster, Cambridge Dictionary, Bilingual Dictionary. It is available in many formats and versions. Mobile dictionary is the most useful and convenient tool for students in the field of education. According to Aslan, large dictionaries were being replaced by portable dictionaries that allow us to access every word we want with less effort. The use of mobile dictionaries is now common for second and foreign language learners around the world. In Nepal language domain in ICT, among all electronic dictionaries, mobile phone dictionaries are widely used for learning and vocabulary development as mobile phones play an important role for everyone including students (Rai, 2020). The mobile app offers users a convenient way to access the dictionary's resources on their smartphones or tablets.

### **Chatbot Technology**

Strategy and logic used for chatbots at Wit.ai is designed to clarify the sentence and the correct answer in various ways. Friends. ai automatically identifies parts that match sentences submitted by the user. These dots can be keywords or concepts, and the neural network is trained to recognize them. Wit.ai can retrieve sentences containing text and even images. The chatbot works with the following languages: Spanish, English and Catalan. Wit.ai works with content and has four types of content: character, keyword, free text, and keyword and free text. The website itself is the only resource this tactic uses for the chatbot. This level serves as a database of places, including cities and locations with language comprehension and proper response (Gadea & Felipe, 2019). The main limitation of the project is the accessibility of the relevant technologies. The systems required to power Artificial Intelligence applications that are usually very expensive. Because of that, their use was generally sold as a service, and free versions have a server request limit that restricts their potential as mainstream applications. Whereas, our proposed system is there to integrating a chatbot into the dictionary platform, users gain an interactive way to explore linguistic resources through natural language conversations. Furthermore, it can be integrated into both the web-based dashboard for administrative support and mobile app for user engagement.

## **THEORETICAL FRAMEWORK**

### **Activity Theory**

Activity Theory, rooted in the work of Vygotsky and later expanded by Engeström, provides a framework for understanding how tools mediate human activities. In the context of digital systems, it emphasizes the interaction between users and the technology as they engage in meaningful activities, such as language learning (Yamagata-Lynch, 2010). The Duramazwi Digital System serves as a

mediating tool within the Activity Theory framework, facilitating language learning and usage among Shona speakers. By integrating various modules such as speech recognition, translation, and contextual rewriting, the system supports users in their language activities, aligning with the core principles of Activity Theory. The system's collaborative features, where users contribute new words and definitions, reflect the community aspect of Activity Theory, where learning and development are socially constructed. Additionally, the system's ability to evolve based on user interactions exemplifies the dynamic nature of the 'activity system' in Activity Theory. Activity Theory informed the design of the Duramazwi Digital System by emphasizing the need for a user-centered approach. The system was designed to adapt to the user's activities, providing relevant tools and feedback that enhance their learning experience.

## **METHODOLOGY**

In this section we lay out and describe the study plan, research methodology, and research limitations in this part. The researchers utilized the Design Science Methodology, as well as a mixed method approach, which is based on logical principles that contribute to the thesis's ultimate objective. It is both quantitative and qualitative in nature.

### **Design Science Methodology**

The existing Shona dictionaries do not have integrated modules for speech recognition, translation, summarization, and paraphrasing. This fragmentation hinders a seamless flow of information processing and limits the user experience, particularly for linguists, educators, and learners of the Shona language. The need for a comprehensive digital solution that integrates these functionalities is therefore critical. A review of existing digital dictionaries and language processing tools reveals that none adequately address the specific needs of Shona language processing. Current tools lack the integration of speech recognition, translation, summarization, and cultural context, which are essential for a comprehensive language learning and processing platform. This gap in the extant IT knowledge-based underscores the need for a novel solution tailored to these requirements. The Duramazwi Digital System is a novel IT artifact that integrates speech recognition, translation, summarization, and paraphrasing into a cohesive platform designed specifically for the Shona language. This system not only provides a comprehensive dictionary but also supports language learning and cultural preservation through advanced NLP and AI techniques. The utility of the Duramazwi Digital System was rigorously evaluated through a series of tests, including user acceptance testing with educators and students, performance benchmarking of the integrated modules, and analysis of user feedback. The development of the Duramazwi Digital System underscores the importance of culturally sensitive IT solutions in preserving and promoting minority languages.

### **Interviews**

Interviews were performed with randomly selected undergraduate students, specifically those specializing in economic education, and the responses were outlined below. When asked how convenient it is to use a digital dictionary, undergraduate students responded that it is very convenient to carry around at all times. The majority of undergraduate students who responded to the question of whether

digital dictionaries could take the place of printed ones said that they could as they were user-friendly and kept up with technological advancements. Furthermore, the digital dictionary may not always accurately comprehend the meaning of the word that they are searching for. When asked about the advantages of using a digital dictionary, the undergraduate students responded that it makes learning Shona easier and allows them to do assignments more quickly, translate challenging words they haven't understood, and improve their Shona pronunciation because there are various kinds of digital dictionaries that also provide pronunciation information (Desnaranti, L, Putra , & Utama, 2022).

## **Discussion**

In the development of a digital dictionary, discussion is an essential method of data collection. It is a qualitative research approach that involves conversing with individuals or groups in order to gather insights, opinions, and experiences related to the topic being studied. This can aid in guaranteeing the impartiality and accuracy of the data gathered. Teachers and students thought that using a mobile digital dictionary was simple and easy, and that they do not need any training to use one.

## **Crawling and Scraping**

Using web crawlers or scraping programs to gather data from internet sources is one popular technique (Michalakidis & Moschoyiannis, 2016). This can include linguistic data kept on webpages, databases, and other digital archives. The Norwegian Project (Østby, 1998) the first team to electrifies dictionaries. There are numerous museums and collections at Norway's four universities that house data on a variety of topics related to Norwegian language and culture followed by the ALLEX team in Harare (Chimhundu & Mberi, 2009). The ALLEX Project was planned as a project in four phases, each of which would result in publication of one or more monolingual dictionaries. Phase 1 was between 1992 to 1995, they started with General Shona Dictionary then from 1996 up to 2000, General Ndebele Dictionary for phase2. Phase 3 and phase 4 was between 2001 to 2009 they managed to come up with scholars of junior level of Shona and Ndebele dictionaries.

## **Crowdsourcing**

Crowdsourcing is a method used by certain digital dictionaries in which users from all over the world contribute definitions, translations, examples, and other language information. Sites such as Wiktionary make considerable use of this technique. We aim to create a multilingual dictionary with comprehensive information about every word in the Shona language. Given the size of each language, the large amount of information and data that can be associated with each word, the number of languages and variants, and the impossibility of many people reporting on a large number of small languages. However, each word sets the goal of designing a system: a format that can receive complete sets of linguistic data on each word, and a system that can transmit and process the data to speakers of any language.

## **SYSTEM ANALYSIS**

Detailed information on the system and its users is provided in this section. Then it shows the functional and non-functional needs that were gathered by employing various techniques like interviewing, surveys, and brainstorming. Following the selection of the most crucial requirements, requirement

analysis was implemented utilizing a variety of methods, including class diagrams, sequence diagrams, and activity diagrams.

### **Requirements analysis**

System requirements analysis is the process by which researchers gain a thorough understanding of the system as defined in the system aim and objectives. The researchers will outline the functional, non-functional, software, and hardware requirements of the Duramazwi Digital in this section.

#### *Functional requirements*

- Speech recognition - convert speech to text, enable voice input
- Question answering / search / dictionary - perform text-based searches, retrieves definitions and meanings
- Rewriting, paraphrasing, summarizing -translation capabilities multilingual support, contextual rewriting.

#### *Non-functional requirements*

- Performance -the digital dictionary should provide fast response times for searches, translations, and definitions.
- Scalability -the system should be capable of handling increasing user demand and accommodating future growth
- Usability -the digital dictionary should have an intuitive user interface, easy navigation, and clear instructions.
- Offline functionality -the system should provide offline access to previously accessed words and definitions.

## **BUSINESS MODELLING PROCESS**

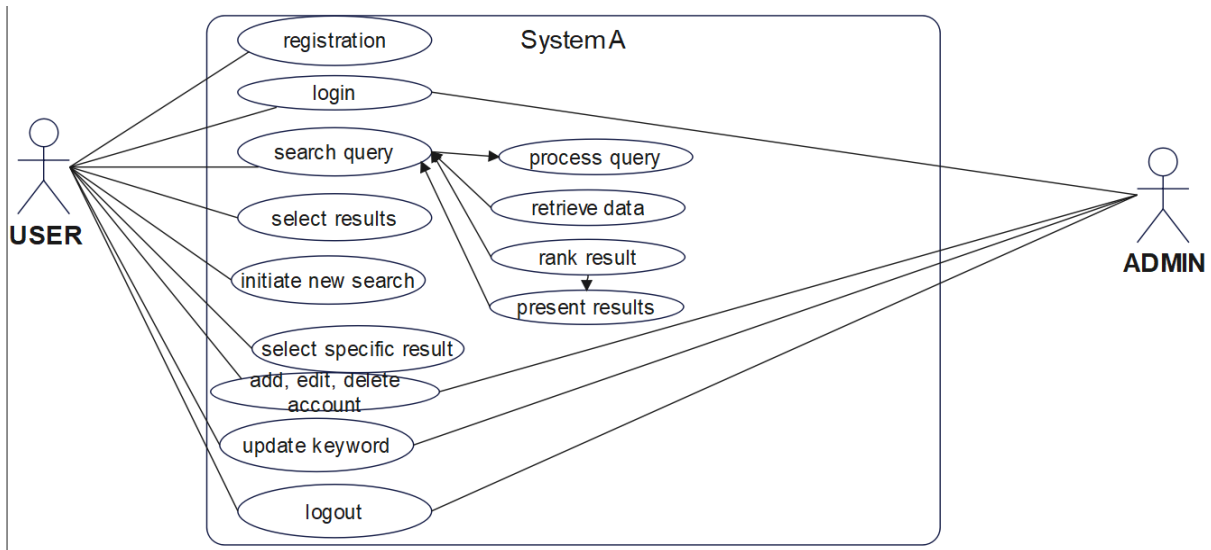
This is a graphical method of illustrating how business processes are related and interconnected through the use of flowcharts and activity diagrams. Business process modelling depicts the events and activities that take place in the business. Services can carry out business activities, so business process modelling depicts the representation of business operations in a graphical and quantitative format.

### **Use case model**

A use case diagram's primary objective is to illustrate the changing nature of a system. The system's requirements, including both internal and external effects, are accumulated by it. It displays the potential interactions between a system component and an external object (Aleryan, 2016).



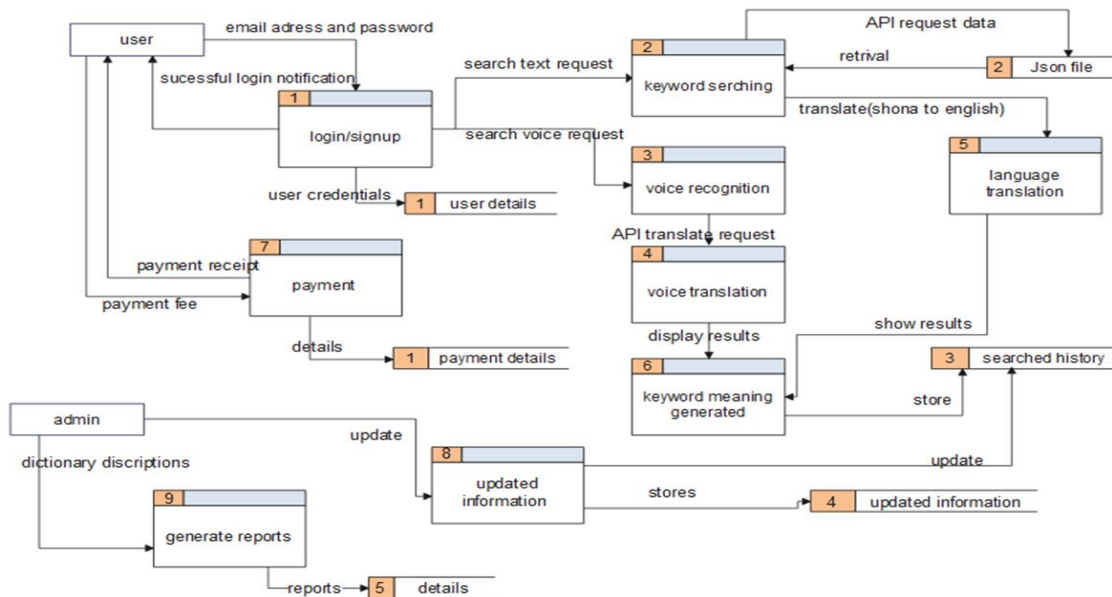
**Figure 1 Use case diagram**



**General subsystem dataflow**

The context diagram is divided into various bubbles and activities in 1-level DFD in figure 2. In this stage, we highlight the system's primary functions and decompose the high-level DFD process into smaller processes. Additionally, Level 1 DFD discusses fundamental procedures and information sources (Ibrahim & Yen, 2010).

**Figure 2 Data flow diagram**



## DURAMAZWI DIGITAL COMPUTER TESTING

It is the process of running a program or application in order to detect software problems. Testing has the objective of verifying and validating of the Duramazwi Digital System for the visually impaired that meets user requirements, in short to see if the system is doing what it is supposed to do.

### Integration testing

The individually tested units were progressively combined into more complex groupings and then tested. These groupings were exhaustively tested and analyzed until the entire system has been put together and tested. The test results produced positive results as the system was functioning properly.

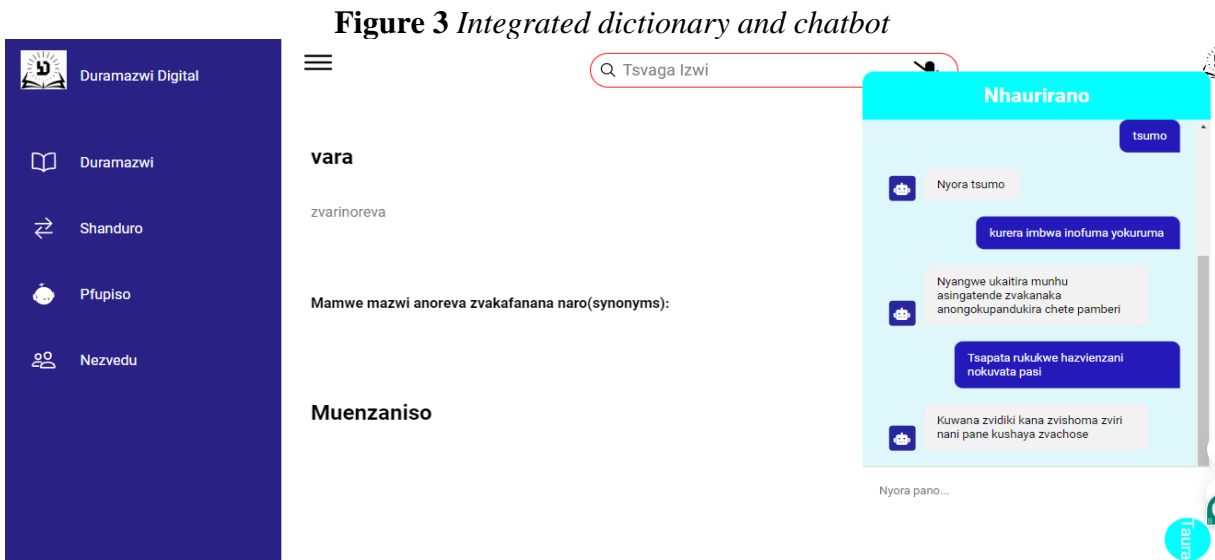


Figure 3 Integrated dictionary and chatbot

The functionalities include how the user interacts with the system, in this case the visually impaired. Also, the different modules were tested separately before being integrated to ensure that there are no bugs. Then after integration the modules are working perfectly, each performing its function well.

### Chatbot (Nhaurirano)

When utilizing a chatbot designed to interpret idioms(*madimikira*) and proverbs(*tsumo*). Idioms and proverbs are expressions or phrases that hold a figurative meaning different from the literal interpretation of the words used. These linguistic elements often carry cultural significance and may not directly translate word-for-word. Therefore, when interacting with our chatbot it is essential to input well-known or commonly used phrases to receive an appropriate response for example an apple does not fall far from the tree (*mbudzi kudya mufenje hufana nyina*). If the chatbot encounters an input that is unclear, it should prompt the user in a diplomatic manner to rephrase or provide additional context for better comprehension. By encouraging users to refine their inputs politely (*zama kunyora futi patsva*), the chatbot can facilitate effective communication and deliver more accurate results.

### Tone marking

Tone marking in Shona is very important as it distinguishes between words spelt the same but pronounced differently. For example, if tone is not marked, there will be no difference between *nzara*

low tone (hunger) and *nzara* low, high tone (finger nail). However, primary school teachers and lecturers consulted during outreach programs argued that tone should not be marked in such a dictionary, as it will confuse children. The dictionary should be very simple and concentrate on the meaning of words.

### Synonyms and variants

Synonyms are lexical items which have the same meaning or are so close in meaning that they can be used interchangeably in certain contexts. In Shona the word baboon can be translated with *bveni*, *dede*, *diro*, *gudo* and *mutiro*, depending on one's dialect. According to Crystal (1997: 408), a variant can be defined as a set of alternatives in a given context. Shona has many variants, mainly because of its different dialects. For example, the *Karanga* people in some instances use *zh* and *gw* where the other dialects like *Zezeru* and *Manyika* use *nz* and *rw* respectively as can be seen in words like *zhara* and *nzara* (hunger), *-gwara* and *-rwara* (to be sick).

**Figure 4** *Synonyms and variants tests*

**amai**

**vara**

zvarinoreva

Izwi rinodaidzwa mubereki wechikadzi

Izwi rokukudza mudzimai wose wose

**Mamwe mazwi anoreva zvakafanana naro(synonyms):**

[mai](#) [mhai](#) [mhamha](#)

### CONCLUSION

By incorporating audio input functionality, users can easily search for words by speaking them aloud, catering to individuals who may find it challenging to type or spell certain words. This feature enhances accessibility and inclusivity within the digital language learning space. Translation capabilities allow users to seamlessly translate words or phrases from Shona to English. This fosters cross-linguistic understanding and promotes multilingualism among users. The availability of idioms in the dictionary adds depth to the learning experience by providing cultural insights and nuances that are integral to mastering a language. The feature that offers synonyms for searched words enriches vocabulary acquisition and aids in expanding linguistic proficiency. By presenting alternative words with similar meanings, users can enhance their expressive capabilities and develop a more nuanced understanding of the Shona language. Overall, a comprehensive Shona dictionary equipped with audio input, translation

features, idioms, and synonyms serve as a valuable tool for language learners, educators, researchers, and anyone interested in exploring the richness of the Shona language.

## REFERENCES

- Aleryan, A. (2016). Comparative Study between Data Flow Diagram and Use Case Diagram. *International Journal of Scientific and Research Publication*, 5.
- Benbada, M., & Benaouda, N. (2023). Investigation of the Role of Artificial Intelligence in Developing Machine Translation Quality. Case Study: Reverso Context and Google Translate translations of Expressive and Descriptive Texts. Language Combination: Arabic-English/ English-Arabic. (*Doctoral dissertation, Faculty of Letters and Languages-Department of English*).
- Benjamin, M. (2021). Crowdsourcing Microdata for Cost-Effective and Reliable Lexicography.
- Chimhundu, H. (2005). Lexicography and Language Raising: Dictionaries in Zimbabwean Languages. *A Public Lecture Delivered as Part of the Golden Jubilee Celebrations of the University of Zimbabwe (13 April 2005)*. Harare: University of Zimbabwe.
- Chimhundu, H., & Chabata, E. (2007). Duramazwi reDudziramutauro noUvaranomwe. *Allex Project African Languages Research Institute*.
- Chimhundu, H., & Mberi, N. E. (2009). *Duramazwi Rechishona*. Harare.
- Desnaranti, L, L., Putra , F. P., & Utama, W. (2022). The Analysis of Digital English Dictionary Use by Non-English Major Undergraduate Students. *Acitya: Journal of Teaching & Education*, 221-224.
- Emmanuel , C. (2006). *An Advanced Dictionary*. Harare.
- Gadea, L., & Felipe. (2019). CHATBOT. *A digital assistant with built-in AI*, 14-25.
- Ibrahim, R., & Yen, S. (2010). Formalization of the data flow diagram rules for consistency check. *International Journal of Software Engineering & Applications*, 1(4) ,17.
- Kerssens, N. (2016). *An Archaeology of Computing's Integration with Every Life*, University of Amsterdam.
- Lonke, D., Ilan, K., & Dzhuranyuk, V. (2022). LEXICAL DATA API. In *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*,. Mannheim: IDS-Verlag.
- Michalakidis, G., & Moschoyiannis, S. (2016). *Appreciation of structured and unstructured content to aid decision making- from Web scraping to ontologies and data dictionaries in healthcare*. University of Surrey.
- Østby, A. (1998). *The Norwegian Documentation Project* .
- Rai, H. K. (2020). *Students' Perception Towards the use of Mobile Dictionary App for Vocabulary Learning*, Tribhuvan University of Nepal.
- Rundell, M., & Stock, P. (1992). The corpus revolution. *English Today*, 21-32.
- Susilana, R., Gema , R., Ardiansah, & Yuyu , W. (2022). Pedagogia Dictionary: Web Application Development. *International Journal of Instruction*, 15(1), 197-218.
- Yamagata-Lynch. (2010). Understanding Cultural Historical Activity Theory. *In: Activity Systems Analysis Methods* . , Springer, Boston, MA.