



RESEARCH ARTICLE

AN ENGLISH TO YORUBA BILINGUAL LEXICON FOR HEALTH SECTORS USING A DIRECT APPROACH TO MACHINE TRANSLATION

Abiola O.B¹, Babalola G.O², Adaramola O.O³

^{1,2}Department of Computer Science, Afe-Babalola University, Ado-Ekiti, Nigeria

³Department of Computer Engineering, Afe-Babalola University, Ado-Ekiti Nigeria

¹adeoyetoyin@yahoo.com; ²gbemmie5@yahoo.com; ³lekesido@yahoo.com

The importance of indigenous languages in the dissemination of information to local people cannot be overemphasized. This paper is focused on the perceived benefits of one of the major indigenous languages in Nigeria (the Yoruba language), in the learning and usage of diseases names in health sectors. Machine translation (MT) has been defined as a subfield of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. This system employed the use of the direct approach to machine translations in building a bilingual lexicon for English diseases names to Yoruba. The tool has been designed uniquely to provide an easy to use platform for medical practitioners. The effectiveness of the system was evaluated both formally and informally and the implementation was carried out using Visual Basic 6.0. This system will be of immense benefits to people in the health sectors.

Keywords—: Bilingual Lexicon, Indigenous languages, Health sectors, Direct Approach

I. INTRODUCTION

The only channel by which human beings abstract reality is language. The importance of communication to man is incalculable especially when daily human interaction is considered. That is, to learn a language is not only reaching out to others but to maintain a variety of social bond, a shared sense of values and communal awareness. Language is a powerful tool of control which forms a large part of the culture of people. [12]

According to [15], the following were said concerning human language: human language is creative as it allows its users to express new ideas, produce and comprehend new sentences. In human language, there is an inbuilt mechanism in which several meaningless sounds can be combined to derive meaningful longer expressions, human language allows its users to talk about things or situations which are not necessarily present during the time of discourse, human language is culturally transmitted so long as children are exposed to the environment where the language is spoken. Human language allows its users the possibility to perceive as well as reflect upon the message that they transmit themselves and there is no limit to which human language can be employed in communication and every aspect of human behaviour or endeavour can be expressed that is to say that human language is versatile. [15].

In Nigeria, there are three major languages: Yorùbá, Igbo and Hausa respectively. The languages spoken in Nigeria are not evenly distributed despite their large numbers. For instance in the South-West part of Nigeria, Yorùbá is largely spoken; Igbo is largely spoken in the South-East part of Nigeria; while in the North-West part of Nigeria, Hausa is largely spoken. Nigeria languages have been categorized on the basis of the population of their speakers and relative importance into major, minor and minority languages. Hausa, Igbo and Yorùbá are the major languages, while languages such as Fula/Fulfude, Kanuri, Efik/Ibibio, Tiv and Izon are also categorized as constituting other class of major languages but smaller than those in the first category. The minor languages include Edo, Nupe, Urhobo, Igala, Idoma, Epira, Itsekiri, to mention but few. While others not included in these minor groups are referred to as the minority languages as affirmed by [2].

The importance of indigenous languages in disseminating information cannot be overemphasized. The dominance of the English language in Nigeria is quite overwhelming and this can be seen in practically all domains: government and administration, education, the media, the judiciary, science and technology to mention but few. High government officials avoid using their languages in official contacts even with their own people for fear of being labelled tribal and parochial. In the National and State Houses of Assembly, English continues to be the language of debate and record in spite of the fact that provision is made in the constitution for the use of the major indigenous Nigerian languages. Our major indigenous languages are slipping away from us because of the various trans-national structural revolutions going on in the world today in the name of globalization [2].

As a matter of fact, it has become a taboo to talk of the role of indigenous languages in the question of education, health monitoring and national development as a whole. The bigger scandal however is that, all along we have been lying to ourselves and to others about the sufficiency of foreign English language in addressing our developmental problems. Amidst all the factors that led to developmental problems in all sectors be it health, education, etc language factor stands out as the most critical. In health sector for instance a medical practitioner from the north posted to a rural area in the south who finds it difficult to communicate with dwellers in such community as a result of language barrier, how will he or she prescribe appropriate drugs for patients who cannot express himself or herself in English language in such environment? A teacher from the east teaching in a rural area in the north where students cannot fully express themselves in English how will such a teacher impact knowledge appropriately? There is need to find a way of appreciating the beauty of our indigenous languages by properly engaging and experimenting them as languages of instruction in schools, hospitals among medical practitioners, business transactions and even in national development so as to complement English language which still serves as the sole vector of knowledge in the country. English as a first language, however, remains an exclusive preserve of a small minority of the country's urban elite, and is not spoken at all in some rural areas.

This paper aim to address the issue of dissemination of information among people in health sectors and dwellers of some rural areas in Yoruba land through the use of computers. It is a platform designed to provide easy to read guide for all diseases names and their various meanings in Yoruba language. It covers all must-know diseases names in English as well as their various descriptions in both English and Yoruba language which we believe will go a long way at helping medical practitioners, and other people in health sectors from the eastern or northern part of the country who find themselves in the southern part to get to learn how to pronounce names of diseases in Yoruba as well as their descriptions so as to improve their work and provide appropriate means of comprehending and disseminating information to rural community especially those with little or no understanding of the English language.

Natural language processing is an area of research that addresses translation from one human language to another and machine translation is an aspect of natural language that is used by researchers to digitalize information from one natural language to another. Developing a Machine Translation(MT) system, that accurately produces a good translation system between human languages has been the target of researchers of Machine Translation (MT). However, despite decades of great efforts to develop machine translations for languages, efficient methods for machine translation continue to be a challenging task. Although, efforts on machine translations from one language to another have been in existence for long, a fully-automatic general purpose high quality machine translation system (FGH-MT) is extremely difficult to build and there is no system in the world of any pair of languages which qualifies to be called FGH-MT as affirmed by [14].

Machine translation mainly deals with the transformation from one natural language to another. Natural language Interface provides the user freedom to interact with the computer in a natural language used for day to day communication. One of the important goals of computational linguistics is a fully automatic machine translation between such natural languages. This is important because communication between people from different linguistic backgrounds still poses as a major problem. [10]. Machine translation of natural languages is a very difficult task. It can be perceived as the simple substitution of words in one natural language for words in another. Yet it is not so simple because of the complexity of some natural languages. Many words have various meanings in some languages and so they can be translated in different ways. Also, sentences might be ambiguous and have various meanings. Grammatical relations can vary depending on the languages, and

translating sentence from languages having different relations means reformulating the sentence. Besides, problems due to the associated world knowledge may be encountered and these are usually difficult to solve [1].

To translate from English to Yoruba using any of the approaches to machine translation, as a result of the differences in their syntax and structures, there is need to have a very good knowledge of the source language, a very good knowledge of the grammatical features of the target language, familiarity with the subject matter of the text being translated and a profound understanding of the etymological and idiomatic correlates between the two languages as well as a finely tuned sense of when to metaphrase ("translate literally") and when to paraphrase so as to assure true rather than spurious equivalents between the source and target-language texts and a profound understanding of the syntax and grammatical features of the two languages as well as their vocabularies. This paper is based on developing a bilingual lexicon or dictionary for names of the various diseases affecting humans from English to Yoruba language and the direct approach to machine translation is employed in the translation process. This platform will be useful to people in health sectors as it will enhance effective communication between medical practitioners and patients especially those in the southern rural environments.

II. THE DIRECT APPROACH TO MACHINE TRANSLATION

The direct approach to translation system, as illustrated in Figure 1 below is a simple and efficient strategy of translation in which the source text goes through analysis and synthesis and with the help of the bilingual lexicon, the target text is produced. It is a method based on dictionary entries, which means that the words will be translated as they are by a dictionary. In this direct approach words are translated directly without passing through an intermediate language construct. The direct translation simply changes single words from one language to another. According to [13], the direct approach is considered to be the most primitive approaches carrying out replacement of words in the source language with words in the target language. This is carried out in the same sequence and without much linguistic analysis or processing. The only resource used by this approach is the bilingual dictionary and that is why it is known as dictionary-driven machine translation. [13].

The direct approach has proved to be very useful in machine translations and it does not require human post-editing in most of the cases. It was chosen for this research because this work involves building a bilingual dictionary and this approach has a lot more power, flexibility and control when translating word for word. It carries out word by word translation with the help of a bilingual dictionary usually followed by some syntactic rearrangement. It is the most basic form of translation that translates individual words in a sentence from one language to another using a two-way dictionary. It makes use of a very simple grammar rules. Direct-based machine translation systems take a monolithic approach towards development. That is, they consider all the details of one language pair. [11], [7], [5]. The direct-based machine translation system starts with morphological analysis which removes morphological inflections from the words to get the root word from the source language words. The next step is bilingual dictionary lookup. A bilingual dictionary is looked up to get the target-language words corresponding to the source-language words. The last step is the syntactic rearrangement in which word order is changed to that which best matches the word order of the target language. This approach to translation is highly robust and it does not break down or stop when it encounters unknown words, unknown grammatical constructs, or ill-formed input. [11], [7], [5].

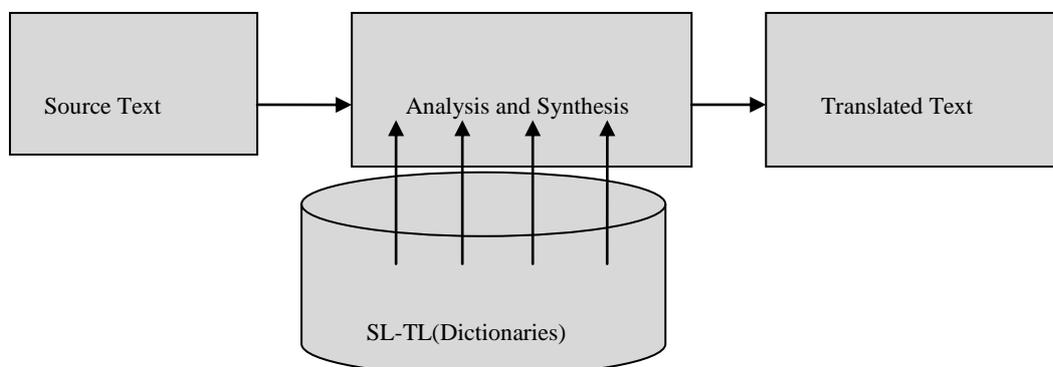


Figure: 1 A Direct Translation System

III.SYSTEM ANALYSIS AND DESIGN

The system includes a comprehensive analysis with the aim of determining the requirements of the system and how these requirements should be satisfied. It involves the architectural design or block diagram as well as the operational flow diagram of the system. The system involves English as the source language and Yoruba as the target language. The system is designed to provide an easy to learn platform for medical practitioners willing to learn names of diseases as well as their descriptions in Yoruba language. It therefore focuses on development of a bilingual lexicon for diseases names in English to Yoruba. Our translation system is a direct translation system since it involves word for word translation in which little or no changes is necessary in word places. The system has five main blocks as shown in figure 2. The process begins with the source language input word block, where a user input disease name in the source language (English) and then the input word goes through pre-processing where the word is stored to check its characteristics. The system stops whenever a space is encountered to signifies the end of a word and automatically eliminate space. The next block is the lexical transfer block which performs the lexical transfer of the word by assigning to it, its corresponding target word counterpart and this is done with the use of the bilingual lexicon which is made up of a bilingual exhaustive lexicon of Yorùbá and English diseases names. In the target language syntactic generator block, the source language word translated to target language counterpart is processed and the output word is produced in the target language. The analysis of the system is further illustrated in the operational flow diagram shown in figure 4 below.

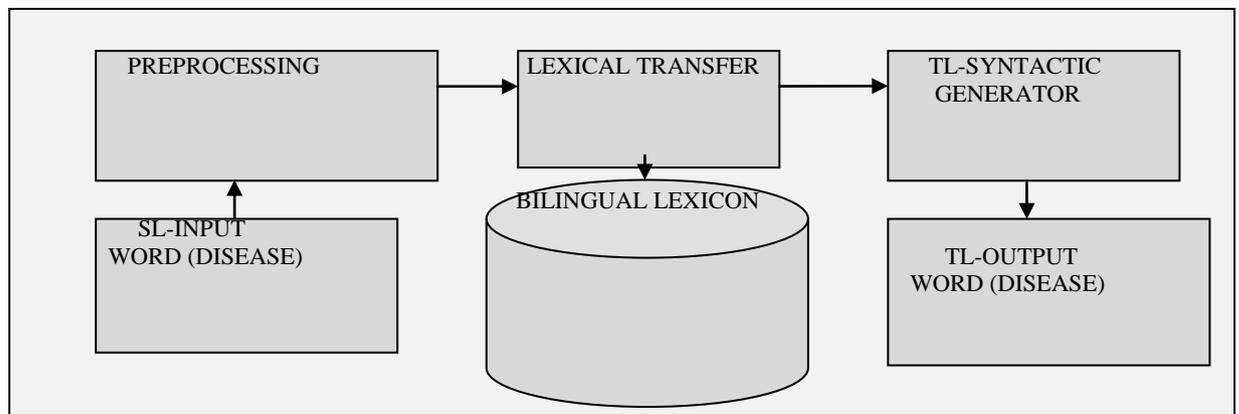


Fig 2: The Translation System Blocks Diagram.

An example to further illustrate the working of the system is shown in figure 3 below: when a user enters the name of a particular disease in English language, say “Cancer” once it enters the next block, it goes through pre-processing which involve analysis and synthesis to obtain the target language equivalent name in Yoruba for the input by obtaining its properties from the developed bilingual lexicon. The equivalent name for “Cancer” in Yoruba language is “Arunjeje” Other examples of some diseases names in our database are illustrated in Table 1.

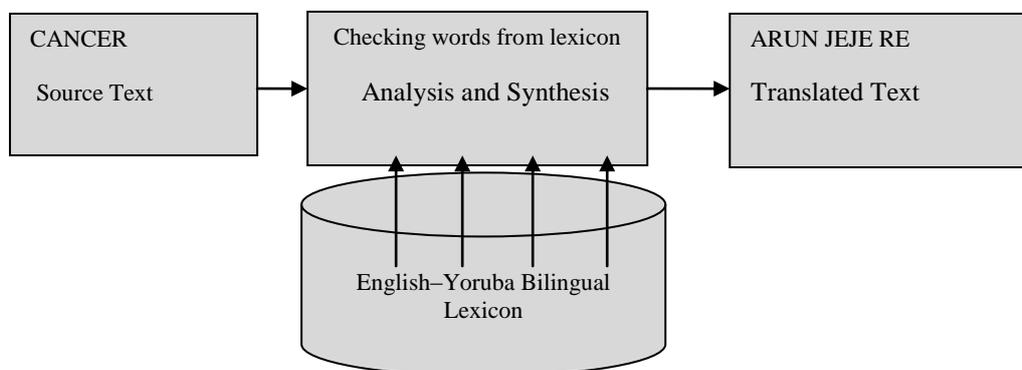


Fig 3: An example of a direct translation.

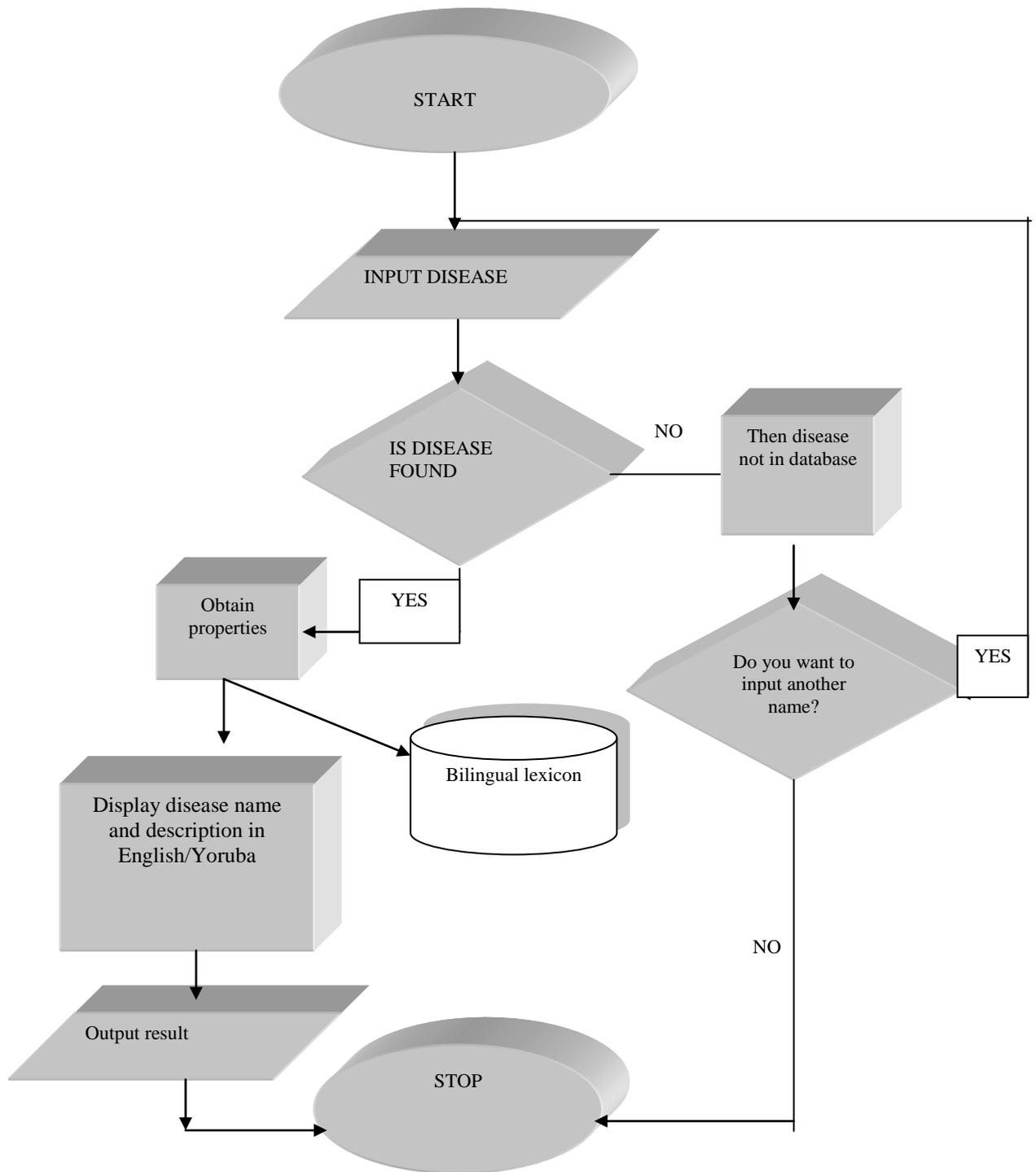


Fig 4: The Operational Flow Diagram.

Table 1: SOME DISEASES NAMES AND THEIR YORUBA EQUIVALENT NAMES

| DISEASES (ENGLISH) | AARUN (YORUBA) |
|--------------------|-------------------------------|
| Malaria | Iba |
| Asthma | Amiso |
| Cancer | Arun jejere |
| Fibroid | Oyun iju |
| Goiter | Gbegbe |
| Cholera | Arun onigba meji or Igbeguuru |
| Tuberculosis | Iko fee |
| Dermatitis eczema | Ifo |

IV.SYSTEM IMLEMENTATION

The system requirements include the technical equipment required for the system to be fully implemented. The requirement is classified into hardware and software respectively. Visual Basic 6.0 was used because it is easy to learn and use and it enables rapid application development (RAD) of graphical user interface (GUI).the hardware requirement include: a minimum of 40GB hard disk, 1GB RAM, Konyin Nigeria Multilingual Keyboard which serve as text editor for Yoruba language. The system is designed as a user friendly system, affording a novice the opportunity to operate the system without difficulty. The tool was designed to provide technical solutions for the usage of the Yoruba language by providing a global, easy to read guide for all diseases names that learners need to communicate with. Below are some interfaces of the system

Welcome Home Page

This page shows links to other pages, it is the master page where all other pages are being displayed. It serves as content-place-holder. It is the only static page in the application where all other pages are dynamic pages.

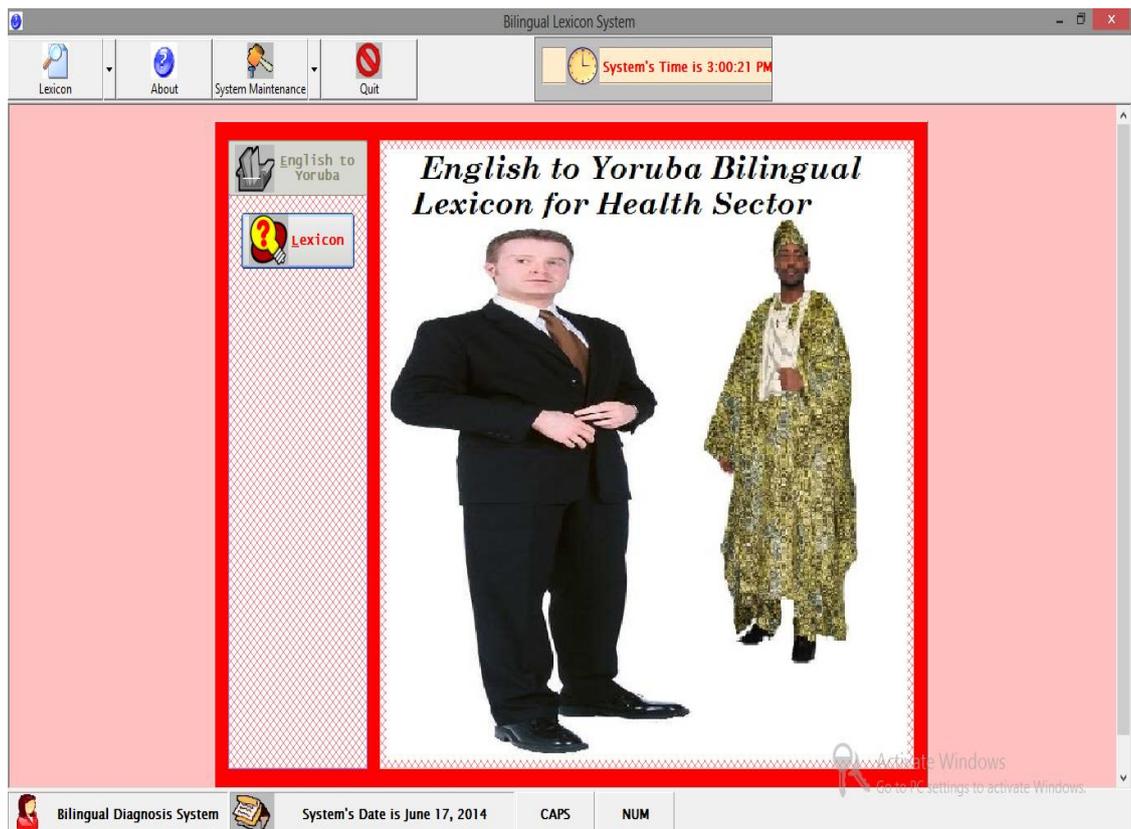


Figure 5: The Home Page

Bilingual-Lexicon Page

This is the searching page where user enter to search for names of diseases and their corresponding meaning in Yoruba language as well as the description of such disease(s) in English and Yoruba. This Interface provide a learning platform for users. From the example in figure 6, “asthma” is typed in the text box and its Yoruba meaning is given as “Amiso” in another textbox. Also, the definition of this disease is provided in another text box as a chronic lung disease that inflames and narrow the airways. The description of the definition is also given in Yoruba.

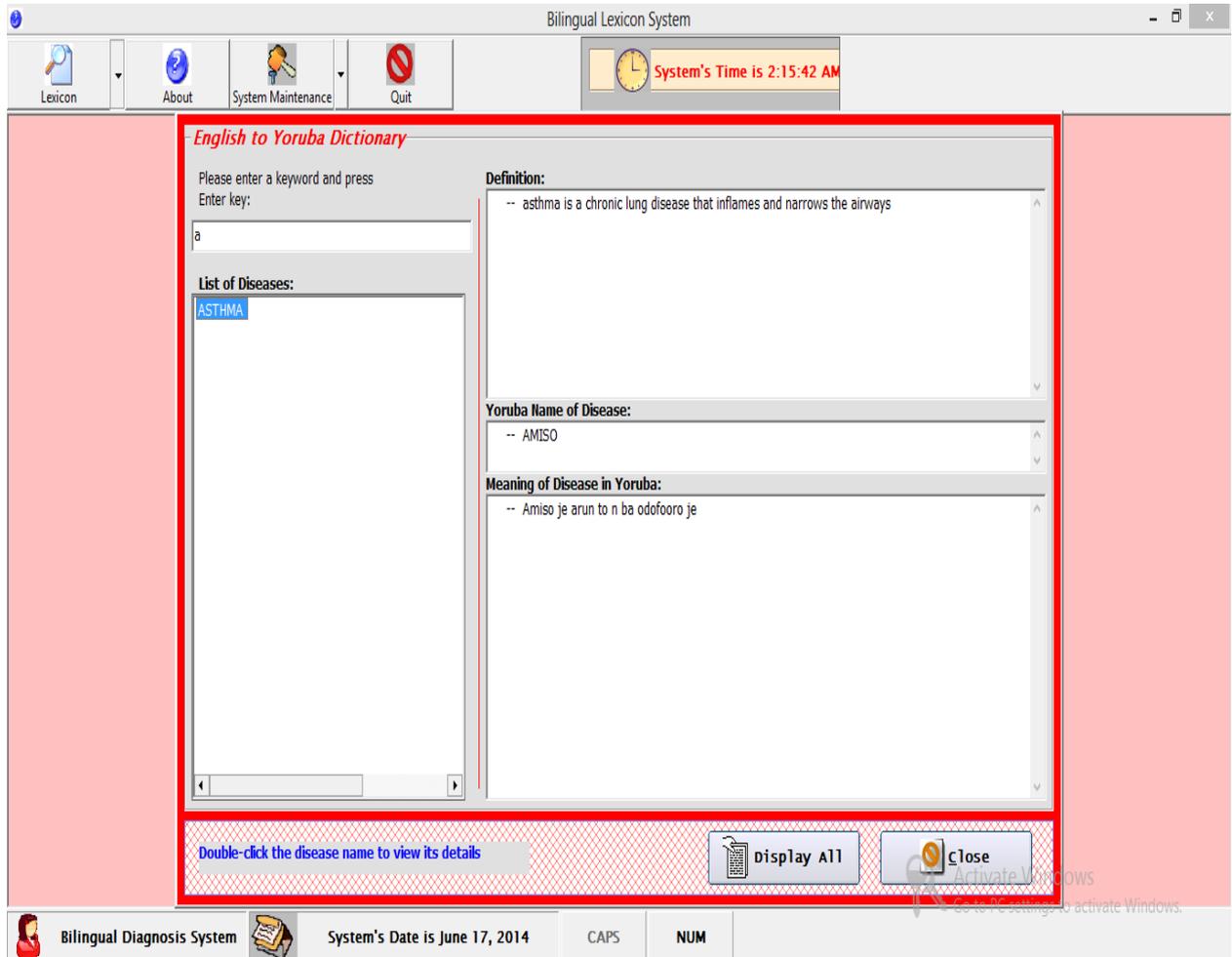


Figure 6: Bilingual-Lexicon Page

Add or Update Interface

This is the Interface where more names of diseases their corresponding meanings, as well as their various descriptions are added to the already existed ones in the database. The bilingual lexicon is updated on this interface.

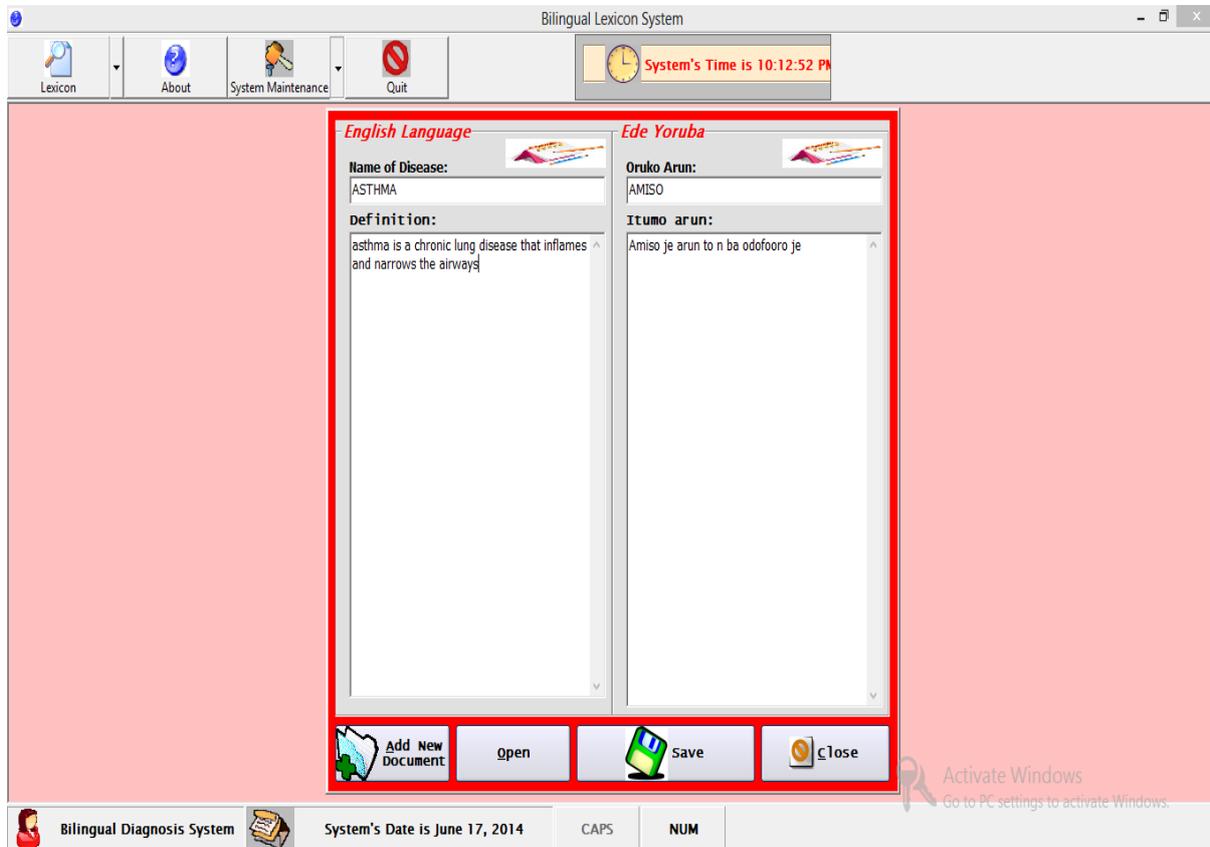


Figure 7: Add or Update Interface

V. CONCLUSION

In this work, we have implemented a bilingual lexicon for English to Yoruba diseases using a direct approach to translation. The current version of our system performs simple substitutions and word for word translations that produces promising and acceptable translations. This system can be improve to achieve a higher quality translation and we believe that with some modifications like multi-word expression transferring module and sentence level post-processor, the system will produce more satisfactory results. However, successful implementation of direct machine translation of two dissimilar languages was demonstrated through this work.

REFERENCES

- [1] Abiola O.B, Adetunmbi A.O, Oguntimilehin A (2015) "Review of the Various Approaches to Text to Text Machine Translations" *International Journal of Computer Applications*. Vol120 No 18, June, 2015, pp 7-12. ISSN: 0975-8887
- [2] Adeoye O.B (2012) *Web-Based English to Yoruba Noun-Phrases Machine Translation System*. A Thesis submitted to the Department of Computer Science, Federal University of Technology Akure, 2012.
- [3] Amarpreet Kaur and Er. Jyoti Rani ; A Review on a Web-Based Punjabi to Hindi Statistical Machine Translation System. *International Journal of Advanced Research in Computer Science and Software Engineering* (IJARCSSE), vol 4, Issue 8, August 2014, pp 383-386, ISSN: 2277 128X.
- [4] Francis M. Tyers (2010) "Rule-Based Breton to French Machine Translation" *European Association for Machine Translation*.http://www.ofisbzh.org/bzh/resources_linguistiques/index-troerofis.php
- [5] Jorg (2009) "Machine Translation: Rule-Based Machine Translation and Machine Translation Evaluation" *Department of Linguistics and Philosophy, Uppsala University*.
- [6] Kamaljeet Kaur batra and G.S Lehal (2010) "Rule Based Machine Translation of Noun-Phrases from Punjabi to English" *International Journal of Computer Science(IJCSI)*, Vol 7, Issue 5, ISSN 1694-0814, pp 409-413. www.ijcsi.org
- [7] Michel Simard, Nicola Ueffing, Pierre Isabelle and Roland Kuhn (2007) "Rule-Based Translation with Statistical Phrase-Based Post-Editing" *Proceedings of the Second Workshop on Statistical Machine Translation*, pp203-206, Prague, June 2007 Association for Computational Linguistics.
- [8] Nakul Sharma (2011) "English to Hindi Statistical Machine Translation System" *Masters Thesis, Tharpa University, Indian*.
- [9] Pankaj Kumar and Er.Vinod Kumar (2013) *Statistical Machine Based Punjabi to English Transliteration System for Proper Nouns* *International Journal of Application or Innovation in Engineering and Management (IJAEM)*. Vol 2, Issue 8, ISSN 2319-4847, pp318-321.
- [10] Sangeetha.J, S. Jothilakshmi, R.N.Devendra Kumar (2014), "An Efficient Machine Translation System for English to Indian Languages Using Hybrid Mechanism" *International Journal of Engineering and Technology (IJET)*, pp 1909-1919, Vol 6 No 4 Aug-Sep 2014, ISSN: 0975-4024.
- [11] Sneha Tripathi and Juran Krishna Sarkhel (2010) "Approaches to Machine Translation" *Annals of Library* Vol 57, pp 388-393.
- [12] Shaibu Sunday Danladi (2013) "Language Policy: Nigeria and the Role of English Language in the 21st Century" *European Scientific Journal*, Volume 9, No 17 (ISSN 1857-7881) Print e-ISSN 1857-7431. pp 1-21.
- [13] Venkateswara Prasad T. and Mayil Muthukumaran.G (2013) "Telugu to English Translation using Direct Machine Translation Approach" *International Journal of Science and Engineering Investigations (IJSEI)*, pp 25-32, vol2 issue 12 January 2013, ISSN:2251-8843.
- [14] Vishal.G. and Gupreet.S.L (2010) "Web-Based Hindi to Punjabi Machine Translation System" *Journal of Emerging Technologies in Web Intelligence*, Vol.2, No 2, May, 2010. Pp 148-151. www.academypublisher.com/ojs/index
- [15] Yusuf et al, (2006) "Basic Linguistics for Nigerian Languages Teachers" *Published by Linguistics Association of Nigeria in collaboration with M and J Grand Orbit Communication Limited; and Emhai Press Port-Harcourt*. ISBN 978-33527-4-2.