

Developing the Ede Language Processing Applications Based on the Vietnamese-Ede Vocabulary Database

Hoang Thi My Le

Abstract— Teaching and learning Ede language programs parallel to Vietnamese language programs have built to support to Ede pupils in learning other subjects, especially Vietnamese. The teacher has explained in Ede language while teaching other subjects to help Ede pupils understanding the content of the lesson deeply and limiting the language linguistic differences between teachers and pupils. So far, Teaching and learning Ede language so far still have difficulties, such as: the number of Ede teachers is still lacking compared to demand; the quality of Ede teaching is still limited due to lack of facilities, equipment, teaching tools and reference documents in teaching and learning Ede language; editing Ede texts with Ede fonts, there are Ede letters that are difficult to remember for typing and easy to cause misspelling; exchange of Ede language documents is limited to displaying Ede writing on computers without Ede fonts; no having Information Technology applications for teachers and Ede pupils in teaching and learning Ede language. To solve a difficult part in teaching and learning Ede language and contribute to improve the quality of teaching and learning Ede language, the paper proposed developing the Ede language processing applications based on the Vietnamese-Ede vocabulary database (V-EBVD).

Keywords— Ede language processing, Vietnamese-Ede machine translation, checking Ede language misspelling

I. INTRODUCTION

In order to apply IT in teaching and learning Ede language, it is necessary to have the efforts of IT experts to invest in developing infrastructure for Ede language processing and building applications in teaching and learning Ede language.

Along with teaching methods, material facilities, standard textbooks and information technology applications, it is also a means to improve the quality of teaching and learning Ede language. In order to create the convenient conditions for teachers and Ede pupils in teaching and learning Ede language, the Ede language processing applications in teaching and learning Ede language has been proposed developing, such as:

Application of looking up Vietnamese-Ede word online and offline, to help pupils and teachers in looking up part of speech, mean of word and the examples clarifying the context of the word. Application of looking up Vietnamese-Ede word offline is also used as a visual aid in teaching Ede language.

Application of checking the misspelling in Ede language

Hoang Thi My Le, Danang College of Technology, The University of Danang, Danang, Vietnam,

documents helps teachers and pupils checking and correcting the syllable misspellings in the Ede documents as lesson plans, books, lessons, exercises...

Application of aiding Vietnamese-Ede machine translation has contributed to translate Vietnamese lessons into Ede language in order to create conditions for further upgrading level of Ede language teaching and help teachers in translating Vietnamese-Ede bilingual lesson plans and contribute to enrich the Vietnamese-Ede bilingual learning through subjects such as: mathematics, science, history, geography ... not only learning Ede language subject.

Application of aiding Vietnamese-Ede machine translation for translating the lessons, lesson plans from Vietnamese into Ede language has contributed to facilitating the further upgrading of the Ede language teaching and learning levels and enriching Vietnamese-Ede bilingual subjects.

II. LOOKING UP THE BILINGUAL WORD IN VIETNAMESE-EDE VOCABULARY DATABASE

A. The role of vocabulary in bilingual teaching and learning

Learning vocabulary and practicing skills of using vocabulary are the first element in teaching and learning a language in general and Ede language in particular.

Because of the importance of vocabulary for Ede pupils, teachers need to develop vocabulary for pupils as well as help them to develop communication skills using vocabulary.

Based on the importance of vocabulary in teaching and learning Ede language, looking up Vietnamese-Ede words is needed for teachers and Ede pupils.

B. Building the application of looking up Vietnamese-Ede word

The restrict needs to be overcome in the building the application of looking up Vietnamese-Vietnamese ethnic minority language word:

- No using Unicode for displaying Vietnamese ethnic minority language writing.
- No using vocabulary database as the infrastructure in Ede language processing.

Overcoming the above restricts, the using Unicode in Ede language text editor and V-EBVD is built [2], [3]. These results used for building the application of looking up Vietnamese-Ede word. The activity diagram is shown in Figure 1.

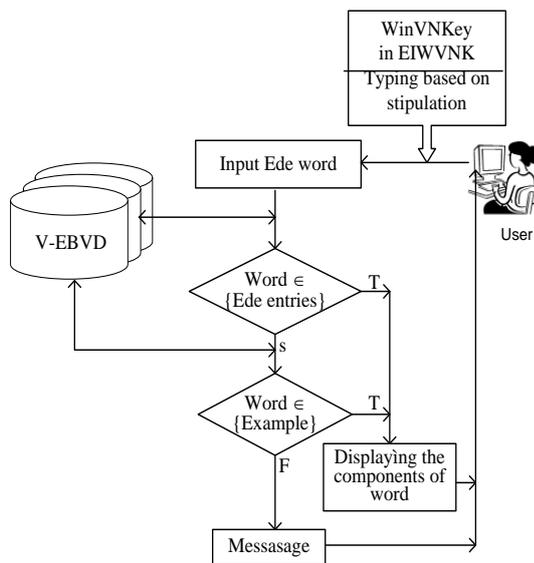


Figure 1. The diagram of the application of looking up Vietnamese-Ede word

1) The application of looking up Vietnamese-Ede word online

The application of looking up online Vietnamese-Ede word has been built in the Ede language processing environment. This environment allows users to look up Vietnamese-Ede or Ede-Viet word.

The Ede letters can be inputted in 2 ways [2]:

- Using EIWVNK
- Input in a way stipulation type

This application has contributed to assist for Ede pupils and teachers in looking up Vietnamese-Ede or Ede-Vietnamese word and its components. The advantages and disadvantages of this application:

- *Advantage:* no installing, everyone using, accessing anytime, updating frequently new entries.
- *Disadvantage:* no anytime also possible to access the internet for looking up word.

Overcoming this difficulty, the application of looking up word offline is proposed to building.

2) The application of looking up Vietnamese-Ede word offline

The database of this application is downloaded from the Ede language processing environment. The Ede letters are inputted in two ways like the application of looking up word online.

The function of displaying vocabulary search results is done through two modules: the module displaying the Vietnamese-Ede and Ede-Vietnamese vocabulary search results:

The module displaying the Ede-Vietnamese search results

Input: WE: Ede word, E: Ede vocabulary database

V: Vietnamese vocabulary, DB: V-EBVD,

Output: {t}: mean of word and its components

Method:

Begin

Input(WE)

Convert(WE)

If Check_Syllable(WE)=False **Then** MsgBox("Error")

Else

If Check_WE(EDE, WE, ED)=1 **Then**

id = Read(E, IDE, WE)

Display(A, DB, IDE=id)

Else

If Check_WE(DB, WE, EXPLE)=1 **Then**

Display({t}, DB, WE ∈ EXPLE)

Else

Msgbox("Not found")

Fi

Fi

End

Figure 2. The algorithm of the module displaying the Ede-Vietnamese search results

The module displaying the Vietnamese-Ede search results

Input: WV: Vietnamese word,

E: Ede vocabulary database

V: Vietnamese vocabulary, DB: V-EBVD,

Output: {t}: mean of word and its components

Method:

Begin

Input(WV)

If Check_WV(VIET, WV, VI)=1 **Then**

id = Read(V, IDV, WV)

Display(A, DB, IDE=id)

Else

If Check_WE(DB, WV, EXPLE)=1 **Then**

Display({t}, DB, WV ∈ EXPLE)

Else

Msgbox("Not found")

Fi

Fi

End

Figure 3. The algorithm of the module displaying the Ede-Vietnamese search results

The computation cost of the algorithm of Figure 2 and Figure 3 depends on the entry quantity of V-EBVD.

This application also used as a visual aid in teaching Ede language and overcome the disadvantages of the application of looking up word online.

III. VIETNAMESE-EDE MACHINE TRANSLATION

To apply machine translating Vietnamese text into Ede language for translating lessons from Vietnamese into Ede language and processing the ambiguity cases that machine translation has not processed. The paper has proposed building the application of aiding Vietnamese-Ede machine translation for translating the lessons, lesson plans from Vietnamese into Ede language.

Application of aiding Vietnamese-Ede machine translation has contributed to translate Vietnamese lessons into Ede language in order to create conditions for further upgrading level of Ede language teaching and help teachers in translating Vietnamese-Ede bilingual lesson plans and contribute to enrich the Vietnamese-Ede bilingual learning through subjects such as mathematics, science, history, geography... not only learning Ede language subject.

A. Solving the Vietnamese word segmentation problem in Vietnamese-Ede translation

The word segmentation problem has not researched. This problem is inherited from vnTokenizer tool that was announced and shared of "theme VLSP" [5], [1].

In order to vnTokenizer tool separates Vietnamese word corresponding with Ede language word, the noted Vietnamese entries is added into expand vocabulary database of vnTokenizer. This adding is necessary for Vietnamese-Ede machine translation.

B. Building the application of aiding Vietnamese-Ede machine translation

1) Translation method

The translation method of this application is based on the V-EBVD. This method depends on the V-EBVD and is almost translated directly by replacing 1-1.

However, only applying this method to translate can't process the ambiguity cases and the words not in vocabulary database. Therefore, the translation result can't be used to translate for lessons in teaching and learning Ede language.

From this restrict, the paper proposed to build the application of aiding Vietnamese-Ede machine translation. The application needs to integrate the supporting of user to identify suitable words for ambiguous cases and words not in vocabulary database.

2) The tool of aiding Vietnamese-Ede machine translation

Operating of Vietnamese-Ede translate aiding application:

Step 1: inputting the Vietnamese text to be translated into Ede language

- Step 2: inheriting form vnTokenizer tool for the segmenting to Vietnamese words
- Step 3: processing the order of interrogative words in Vietnamese questions
- Step 4: translating the segmentation Vietnamese words into Ede language words based on the V-EBVD. The ambiguous and no in V-EBVD Ede language words are highlighted to support for the aiding of user.
- Step 5: processing the order of the "snäk" adjunct in Ede language sentences
- Step 6: integrating the aiding of user for processing the ambiguous and no in V-EBVD Ede language words.

Operating of Vietnamese-Ede translate aiding tool is shown in figure 4.

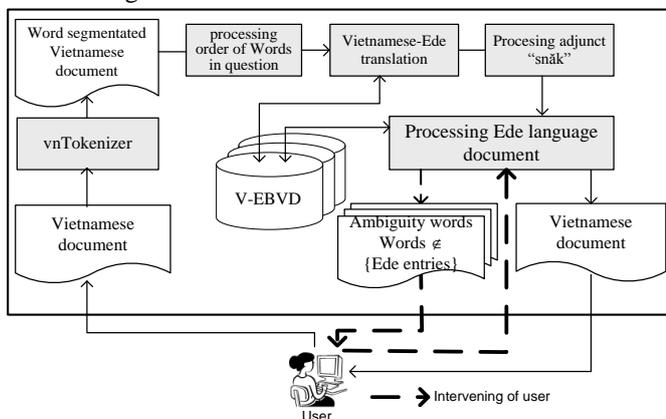


Figure 4. Operating of Vietnamese-Ede translating aiding tool

3) Testing results

From the obtained results, it was found that the interrogative words were not processed due to the user inputting wrong. Therefore, the paper also has the basis to

confirm the order processing module of the interrogative words in the Vietnamese document that can be used in the application of aiding Vietnamese-Ede machine translation.

The order of "snäk" adjunct in sentence has been almost processed entirely. A few cases can't be processed because the next word is not adjective or user inputting wrong.

The tool of aiding Vietnamese-Ede machine translation has contributed to aid for translating lessons from Vietnamese into Ede language and helping teachers in the preparation of textbooks and Vietnamese-Ede bilingual lessons.

IV. CHECKING EDE LANGUAGE MISSPELLING

The typing wrong is the misspelling occurring on documents. This error is more likely to occur on documents inputted by people who are not fluent in Ede language or not familiar with stipulation typing of Ede letters.

In the Ede language processing, The problem of the checking misspelling in Ede document has so far not been announced. From the result of the checking Ede misspelling based on Ede syllable models, the paper continues to develop application of misspelling checking in Ede language text.

A. Solving the problem of the checking misspelling in Ede language text

The problem of the misspelling checking in Ede language document is solved based on the script:

Input: Ede language document, V-EBVD, Ede syllable models

Output: the Ede document is checked misspelling.

Method: the sequence of steps is shown in Figure 5.

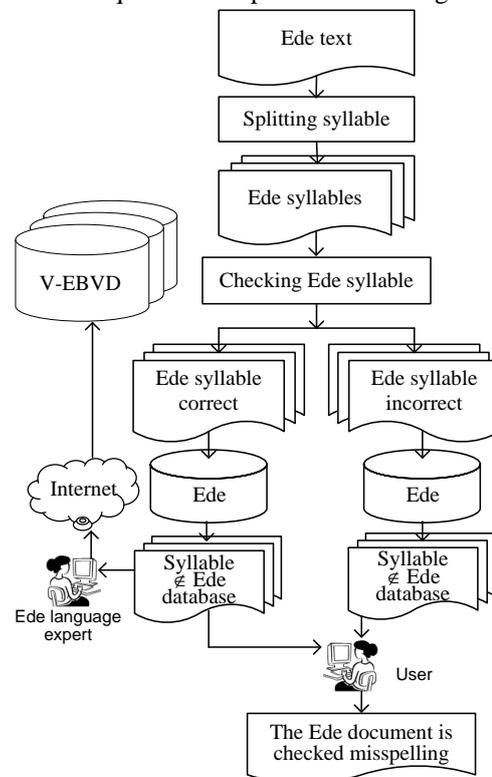


Figure 5. The model of the checking misspelling in Ede language text

B. Building the tool of the checking misspelling in Ede language text

Based on the model of the checking misspelling in Ede

language text [4], the paper proposed building the application of checking misspelling in Ede language document, named SCET (Spelling Checking of Ede Text), shown in Figure 6.

1) Testing installation

The installing testing of the SCET toolkit towards the Ede syllable approach was proposed and obtained testing results were satisfactory. The testing data was Ede language books, Ede-Vietnamese bilingual stories, ethnic and mountainous newspapers. The testing dataset consists of 30 document files with a random misspelling that have been identified, with a total of 150 misspellings.

The obtained results, the SCET found 142 wrong inserted misspellings. The 8 syllables did not find because of the ambiguity between the first part, the main sound and the final sound of syllable. Table I shows the obtained results of the SCET.

TABLE I. Testing Results With Syllables to Be Put Into Document

Inserted misspellings	Finding misspellings	No finding misspellings	Percentage
150	142	8	5,7%

2) Evaluating the testing result

The ability finding Ede syllable misspellings of the solution achieved 94.3% (Table I).

The SCET has contributed to help teachers, pupils finding and correcting syllable misspelling in Ede document as lesson plans, lessons, exercises...

V. CONCLUSION

The solution developing the Ede language processing applications in Ede language teaching and learning Ede language based on the Vietnamese-Ede vocabulary database has contributed to improve the quality of teaching and learning Ede language with the conservation and sustainable development Ede language in the integration trend.

REFERENCES

[1] D. Dien, et al (2006). "A maximum entropy approach for Vietnamese word segmentation", Proceeding of 4th RIVF Vietnam, pp.12-16.
 [2] H. T. M. Le, V. Souksan, P. H. Khanh, (2013), "Using Unicode in Encoding the Vietnamese Ethnic Minority Languages. Applying for the EDe Language", Proceeding of the International Conference on Knowledge and System Engineering, Springer, KSE 2013, Hanoi, pp. 137-148.
 [3] H. T. M. Le, P. H. Khanh (2014), "A Domain indicating method for Ede terminology in building a Vietnamese-Ede bilingual corpus", Proceeding of The third Asian Conference on Information Systems, ACIS'2014, Nhatrang, pp. 434-439.
 [4] H. T. M. Le, P. H. Khanh, (2016), "Basing on the Ede syllable models to check Ede syllable misspelling, applying to improve the quality of Ede vocabulary corpus", Proceeding of the International Conference on Advanced Technologies for Communications, ATC 2016, Hanoi, pp 158-162.
 [5] H. T. Bao, "Building Basic Resources and Tools for Vietnamese Language and Speech Processing" theme, <http://vlsp.vietlp.org:8080/demo/?page=home/>

First Author <http://scv.udn.vn/html>.

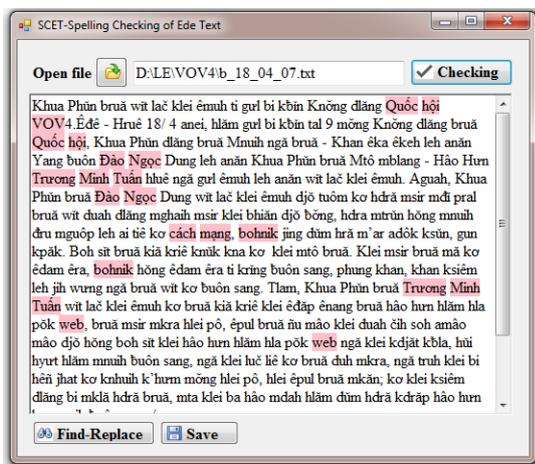


Figure 6. the tool of the checking misspelling in Ede language text