

Proposing a Vietnamese-Hre-Co Interaction Model to Support Building Bilingual Vocabulary Database of Vietnamese-Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese

Le Hoang Thi My
University of Technology and Education
The University of Danang
Danang, Vietnam

Abstract— In order to preserve and promote the national cultural identity; helps ethnic minority students easily acquire knowledge when studying in schools and other educational institutions. Building a bilingual policy for ethnic minorities is really necessary, in order to create conditions for ethnic minorities to learn the spoken and written language of their ethnic group. From the need for a bilingual policy for ethnic minorities, the article has proposed a solution to build a bilingual vocabulary database of Vietnamese-Hre, Hre- Vietnamese, Vietnamese-Co, Co- Vietnamese based on the similarity model. Viet-Hre-Co collaboration, to contribute to creating an interactive environment between users and the vocabulary database. Bilingual vocabulary databases built from the Hre- Vietnamese and Co-Vietnamese interactive model serve as the infrastructure for deploying bilingual vocabulary lookup applications Vietnamese-Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese.

Keywords— Bilingual vocabulary database; ethnic minority; Hre ethnic group; Co ethnic group; interaction model

I. INTRODUCTION

The effectiveness of political, economic, and cultural activities of Party and State officials in ethnic minorities is improved when the speaker directly informs the listener about what needs to be done. Each ethnic minority citizen being proficient in two languages, writing and reading fluently in the national language and ethnic script is a great advantage in the process of improving cultural and scientific levels and expanding relationships between ethnic groups.

In parallel with the bilingual policy for ethnic minorities, in the field of Information Technology, building bilingual Vietnamese-ethnic minority vocabulary databases in general and Vietnamese-Hre, Hre-Viet, Viet-Co, Co-Vietnamese bilingual vocabulary databases in particular is always a challenge.

Starting from the above situation, the article proposes a solution to build a bilingual vocabulary database Vietnamese-Hre, Hre-Viet, Viet-Co, Co-Viet based on the Viet-Hre-Co interaction model, in order to contributing to overcoming the limitations of the existing Vietnamese-Hre, Hre-Viet, Viet-Co, Co-Vietnamese bilingual vocabulary databases. Bilingual vocabulary databases built from the Vietnamese-Hre-Co

interactive model serve as the infrastructure for deploying bilingual vocabulary lookup applications Vietnamese-Hre, Hre-Viet, Viet-Co, Co-Vietnamese.

The structure of the article includes the following contents: First of all, we present the Vietnamese-Hre, Hre-Viet, Viet-Co, Co-Vietnamese. Second, propose a Vietnamese-Hre-Co interaction model in building a bilingual vocabulary database Vietnamese-Hre, Hre-Viet, Viet-Co, Co-Vietnamese. Third, the results achieved and finally the conclusion.

II. STRUCTURE OF BILINGUAL VOCABULARY DATABASE

A. Data criteria

With the goal of building bilingual vocabulary databases Vietnamese-Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese as the infrastructure for deploying bilingual vocabulary lookup applications Vietnamese- Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese. The data criteria set forth in the bilingual vocabulary database are as follows:

- Hre words are mainly collected and recorded in Hre language, which is considered the easiest to hear and understand. Hre language entries partly reflect the traditional culture of the Hre people and are written in Hre script.
- Co language words are mainly collected and recorded in Co language, which is considered the easiest to hear and understand. Items from the Co language partly reflect the traditional culture of the Co people and are written in Co.
- Examples are included to clarify the meaning and usage of words, also known as the context of the entry.
- The word items are labeled with word types: label N for nouns, label V for verbs, label A for adjectives, label O for items that are not nouns, verbs or adjectives.
- Polysemous words are recorded, translated and compared with different equivalent words in the target language.
- When aligning words from the source language, find equivalent words in the target language, based on the basic meaning, commonly used meaning today in both languages.
- Data is stored on the computer in standard Unicode fonts. Maintaining the Integrity of the Specifications

B. Data sources

The Hre-Vietnamese dictionary includes 3,020 word entries, most of which belong to the basic, common vocabulary of Vietnamese. The main vocabulary items are mainly words from Hre language training materials [5], [6].

Co-Vietnamese dictionary includes 1,042 word entries, most of which belong to the basic, common vocabulary of Vietnamese. The main vocabulary items are mainly words from Co language lesson materials [4].

Vietnamese vocabulary with over 31,000 entries, inherited from "VLSP Topic"

Paper dictionary data sources are entered manually, according to a common and consistent form using standard Unicode fonts.

C. Bilingual vocabulary database structure

Bilingual vocabulary database is designed according to the relational database model. Relational databases are used as collections of tables to store data. Database tables are similar to a bilingual vocabulary database, stored completely independently of structure and data. The relational bilingual vocabulary database model has the following advantages and disadvantages:

Table 1, designed to store Hre language entries. NOTE attribute, marks items borrowed from another language.

Table 1: Stored Hre language entries

Field name	Data type	Description
IDH	AutoNumber	Hre index
<u>HRE</u>	Text	Hre entry
NOTE	Boolean	Note

Table 2, designed to store Co language entries. NOTE attribute, marks items borrowed from another language.

Table 2: Stored Co language entries

Field name	Data type	Description
IDH	AutoNumber	Co index
<u>CO</u>	Text	Co entry
NOTE	Boolean	Note

Table 3, designed to store Vietnamese word entries, has about 31,000 word entries inherited from the bilingual vocabulary database in the "VLSP Project" [2]. The ADD attribute to mark Vietnamese entries corresponding to Hre or Co languages is added to the bilingual vocabulary database.

Table 3: Stored Vietnamese language entries

Field name	Data type	Description
IDV	AutoNumber	Vietnamese index
<u>VI</u>	Text	Việt entry
ADD	Boolean	Add

Table 4 is designed to store the results of Vietnamese index and Hre index when interacting according to the Hre-Vietnamese interaction model with data on Viet word index, Hre word index, word types, and other words. For example, Hre- Vietnamese follows the Vietnamese entry.

Table 4: Vietnamese-Hre vocabulary database

Field name	Data type	Description
IDV	Number	Vietnamese index
IDH	Number	Hre entry
SP	Text	Word type
EXPLE	Text	Vietnamese -Hre example

Table 5 is designed to store the results of Vietnamese index and Co index when interacting according to the Co-Vietnamese interaction model with data on Viet word index, Co word index, word types, and other words. For example, Co-Vietnamese follows the Vietnamese entry.

Table 5: Vietnamese-Co vocabulary database

Field name	Data type	Description
IDV	Number	Vietnamese index
IDC	Number	Co entry
SP	Text	Word type
EXPLE	Text	Vietnamese -Co example

III. VIETNAMESE-HRE-CO INTERACTION MODEL

The interactive model is an environment for users to interactively convert paper dictionary data to computer dictionaries online[1], [3].

A. Vietnamese-Hre, Vietnamese-Co interaction model activities

Based on actual Hre and Co language data sources and with the goal of building infrastructure for Hre and Co language processing. The Vietnamese-Hre, Vietnamese-Co interaction model in developing the bilingual vocabulary database Viet-Hre, Hre-Vietnamese, Vietnamese-Co, Co- Vietnamese is proposed, shown in Figure 1.

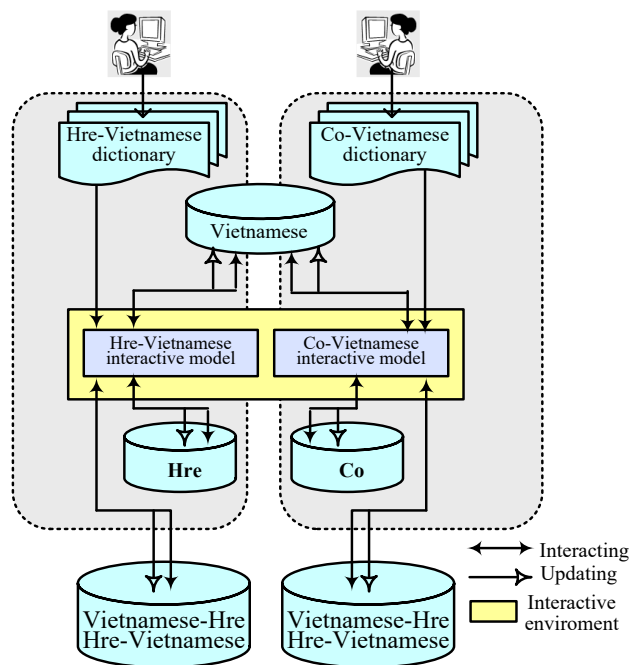


Figure 1. Viet-Hre-Co interaction model

Operations in the model:

- The Hre-Vietnamese and Co-Vietnamese dictionary data sources are formatted before being transferred into the interactive environment.
- The interactive environment reads data from the input data source and interacts with the Vietnamese, Hre and Co vocabulary databases, checking for word entries in the vocabulary databases. or not to perform the update and read the corresponding index. With the Vietnamese index and Hre index, Co index, the environment continues to interact with the bilingual vocabulary databases Vietnamese-Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese. From there, update the bilingual vocabulary database Vietnamese-Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese.
- The interactive environment in the model is built through two interactive modules:
 - The Hre-Vietnamese interactive module has an input data source that is the Hre-Vietnamese dictionary text file.
 - The Co-Vietnamese interactive module has an input data source that is the Co-Vietnamese dictionary text file.

B. Operation of two module

1) Hre-Vietnamese interactive module

Input data sources are Hre-Vietnamese dictionary files, Vietnamese vocabulary database, Hre vocabulary database. The output of this module is the Vietnamese vocabulary database, Hre language database, Vietnamese-Hre vocabulary database and Hre-Vietnamese vocabulary database with updated data.

Procedure of execution

Step 1: read data (Hre words, sets of Vietnamese words, word types and Hre:Vietnamese examples) on each row in the Hre-Vietnamese dictionary file.

Step 2: check Hre words in the Hre vocabulary database, if not there, add them.

Step 3: read the index of the Hre word

Step 4: separate the Vietnamese word from the set of readable Vietnamese words in the second column of the row in the dictionary file. Do this for each separable word in turn:

Step 4.1: check the separated Vietnamese words in the Vietnamese vocabulary database, if not there, add them and make notes to identify new words added to the Vietnamese vocabulary database.

Step 4.2: read the index of Vietnamese words

Step 4.3: extract from the example set the examples Hre : Vietnamese corresponding to the Vietnamese word separated in step 4 in the example set read in step 1. Convert the example Hre : Vietnamese into Vietnamese : Hre.

Example: “Ih oi aip uh?: Ông có khỏe không?”. converted into “Ông có khỏe không?: Ih oi aip uh?”

Step 4.4: check the triple value (Vietnamese word index, Hre word index, word class) in the Vietnamese-Hre vocabulary database.

If not, add the three values and examples of Vietnamese:Hre obtained from step 4.3 to the Vietnamese-Hre, Hre-Vietnamese vocabulary database.

If it exists, check the examples extracted in step 4.3 in the set of examples corresponding to the three value sets. If any examples do not exist, add them.

Step 5: return to step 1 to read data on each row in the Hre-Vietnamese dictionary file one by one until the end of the file.

2) Co-Vietnamese interactive module

Input data sources are Co-Vietnamese dictionary files, Vietnamese vocabulary database, Co vocabulary database. The output of this module is the Vietnamese vocabulary database, Co language database, Vietnamese-Co vocabulary database and Co-Vietnamese vocabulary database with updated data.

Procedure of execution

Step 1: read data (Co words, sets of Vietnamese words, word types and Co:Vietnamese examples) on each row in the Co-Vietnamese dictionary file.

Step 2: check Co words in the Co vocabulary database, if not there, add them.

Step 3: read the index of the Co word

Step 4: separate the Vietnamese word from the set of readable Vietnamese words in the second column of the row in the dictionary file. Do this for each separable word in turn:

Step 4.1: check the separated Vietnamese words in the Vietnamese vocabulary database, if not there, add them and make notes to identify new words added to the Vietnamese vocabulary database.

Step 4.2: read the index of Vietnamese words

Step 4.3: extract from the example set the examples Hre : Vietnamese corresponding to the Vietnamese word separated

in step 4 in the example set read in step 1. Convert the example Co : Vietnamese into Vietnamese : Co.

Example: “Mó kái emănh đhăih?: Ông bà khỏe không ạ?”. converted into “Ông bà khỏe không ạ?: Mó kái emănh đhăih?”

Step 4.4: check the triple value (Vietnamese word index, Co word index, word class) in the Vietnamese-Co vocabulary database.

If not, add the three values and examples of Vietnamese:Co obtained from step 4.3 to the Vietnamese-Co, Co-Vietnamese vocabulary database.

If it exists, check the examples extracted in step 4.3 in the set of examples corresponding to the three value sets. If any examples do not exist, add them.

Step 5: return to step 1 to read data on each row in the Hre-Vietnamese dictionary file one by one until the end of the file.

C. Build an interactive environment

The MVC (Model - View - Controller) design model was chosen to build and manage a bilingual vocabulary database.

The MVC model is a software architecture model used in software engineering. The MVC model divides the application into three design parts, with the goal of separating the interface and code for ease of management, development and maintenance. The design parts of the MVC model include:

The model is responsible for manipulating the database, containing all functions and data query methods such as select, insert, update, delete in the database. The Controller will use those functions and methods to get data and send it to the View.

View is responsible for displaying information to the user through the interface. Display data is received from the Controller.

The Controller receives requests from the user and retrieves the corresponding data from the Model and sends the data through the View to process and return results to the user.

The MVC pattern in action is illustrated in Figure 2.

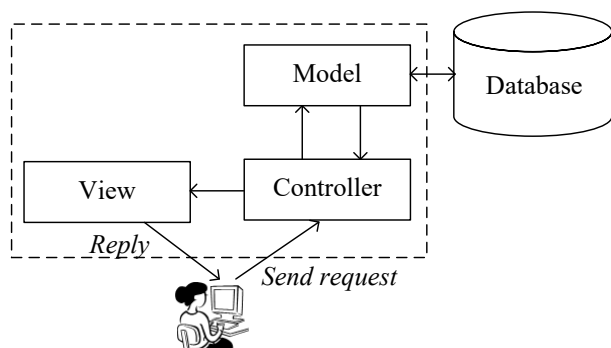


Figure 2: MVC design model

The interactive environment is implemented with two interactive modules to import data from Hre-Vietnamese dictionaries and Co-Vietnamese dictionaries to update into vocabulary databases.

Results achieved

Through the interactive environment, the results of updating the Vietnamese-Hre and Hre-Vietnamese bilingual vocabulary database are listed in Table 6.

Table 6: Statistics on the number of word entries in the vocabulary database.

Vocabulary database	Entries
	Hre-Vietnamese interaction
Hre	3.020
Hre-Vietnamese	3.971
Vietnamese -Hre	4.345

Through the interactive environment, the results of updating the Vietnamese-Co and Co-Vietnamese bilingual vocabulary database are listed in Table 7.

Table 7: Statistics on the number of word entries in the vocabulary database

Vocabulary database	Entries
	Co-Vietnamese interaction
Co	1.042
Co-Vietnamese	1.957
Vietnamese -Co	2.112

IV. CONCLUSION

In the context of processing ethnic minority languages in general and Hre language and Co language in particular, the article proposes to build a bilingual vocabulary of Vietnamese-Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese is based on the Hre-Vietnamese, Co-Vietnamese interaction model, achieving the following results:

- Unified use of Unicode in bilingual vocabulary database.
- Contribute to the development of infrastructure for the problem of processing Hre and Co languages in particular and Vietnamese ethnic minority languages in general.
- Sharing skills for research activities related to processing Hre and Co languages.

The Hre-Vietnamese, Co-Vietnamese interaction model can be extended for the development of Vietnamese-ethnic minority bilingual vocabulary databases.

The proposed solution is practical, because it has contributed to overcoming limitations in the bilingual vocabulary database of Vietnamese-Hre, Hre-Vietnamese, Vietnamese-Co, Co-Vietnamese that the Previous research has not been performed.

ACKNOWLEDGMENT

This research is supported by the Provincial Technology Development and Application Research Project, code 03/2023/HD-ĐTKHCN

Thank for all authors of the papers quoted.

REFERENCES

- [1] Hoang Thi My Le Phan Huy Khanh: "The solution to build a Vietnamese-Ede bilingual vocabulary database based on the Vietnamese-Ede interaction model", No 5 (2), pp. 36–40, 2017.
- [2] Ho Tu Bao, "VLSP topic - Document processing branch" [Online], <http://vlsp.hpda.vn:8080/demo/>
- [3] Le Hoang Thi My, Khanh Phan Huy; "Deploying environment for processing Ede ethnic minority language in Vietnam", IEEE International Conference on System Science and Engineering (ICSSE), 2017.
- [4] Nguyen Minh Tri, "Co language lessons", 2016.
- [5] People's Committee of Quang Ngai province, "Hre language training and fostering materials", 2019.
- [6] Saigon Linguistic Research Institute, "Hre language lessons", Saigon Publishing Ministry, 1974.