http://dx.doi.org/10.5573/IEIESPC.2016.5.2.077

Multilingual Automatic Translation Based on UNL: A Case Study for the Vietnamese Language

Phan Thi Le Thuyen[1] and Vo Trung Hung[2]

[1] Quangtrung University, Binhdinh, Vietnam thuyenptl@gmail.com
[2] University of Science and Technology, The University of Danang, Danang, Vietnam vthung@dut.udn.vn

* Corresponding Author: Vo Trung Hung

Received February 17, 2016; Accepted April 15, 2016; Published April 30, 2016

* *Extended from a Conference: Preliminary results of this paper were presented at the ICEIC 2016. This present paper has been accepted by the editorial board through the regular reviewing process that confirms the original contribution.*

Abstract: In the field of natural language processing, Universal Networking Language (UNL) has been used by various researchers as an inter-lingual approach to automatic machine translation. The UNL system consists of two main components, namely, EnConverter for converting text from a source language to UNL, and DeConverter for converting from UNL to a target language. Currently, many projects are researching how to apply UNL to different languages. In this paper, we introduce the tools that are UNL's applications and discuss how to reuse them to encode a Vietnamese sentence into UNL expressions and decode UNL expressions into a Vietnamese sentence. The testing was done with about 1,000 Vietnamese sentences (a dictionary that includes 4573 entries and 3161 rules). In addition, we compare the proportion of sentences translated based on a direct method (Google Translator) and another one based on UNL.

Keywords: Universal networking language (UNL), UNL system, IAN, EnCo, EUGENE, DeCo

1. Introduction

English is the main language of the Internet, but not everyone knows English. Conversely, a lot of information is written in different languages that are used by only a few people. Knowledge and information are scattered all over the world and remain mostly inaccessible due to non-machine representation and language barriers. Translation is the only means to disseminate such information, but only with great effort involving direct and indirect costs. Most translation involves two languages. To avoid the language barrier, one of the solutions often used is to provide multilingual software or a website written in multi-languages. However, the cost for multilingualization will increase greatly.

An effective solution is to automatically translate into a language that users can read and understand. However, there are more than 5,000 languages, so the development of automatic translation software for each language pair is extremely difficult, especially for languages that have fewer users [2]. So, one of the solutions to translating text from the source language into the target language is to use an intermediate language and develop automatic translation software from any language into the intermediate language. Thus, with n languages, instead of having n*(n-1) translation pairs, we only need 2*n translation pairs [6].

Vietnam has 54 ethnic groups, and people speak in their dialect. The development of automatic translation for all dialect pairs is extremely difficult (2862 pairs). Therefore, the need to find another solution to developing translation services for all of them is necessary.

Universal Networking Language (UNL) is one of the intermediate languages, and at present, translation through

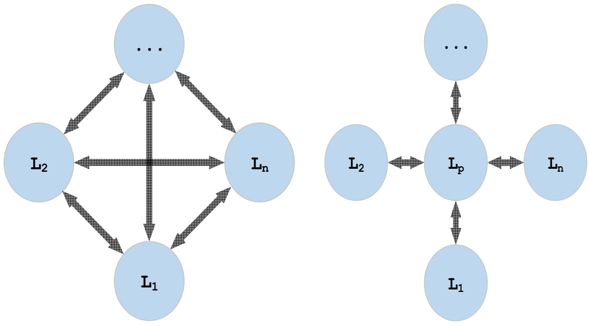


Fig. 1. Models of n*(n-1) or 2*n translation pairs.

UNL has been developed for about 50 different languages [3]. For example, research results for Russian and English are at www.unl.ru; UNL platform tools for 48 languages are at www.undl.org/unlpf, www.unl.ru, www.undl.org/unlexp, and www.eolss.net, etc. In addition, there are a lot of tools for supporting the encoding and decoding process of languages provided by the UNDL at www.undl.org/ and at www.unlweb.net/wiki/tools.

Although UNL has been studied and widely used in the world. In Vietnam, it has not been studied and disseminated. Our team has been carrying out research into UNL since 2005 and has achieved some results, such as studying possibilities to apply UNL to Vietnamese [6], building a UNL - Vietnamese dictionary [3, 4], and testing tools available for UNL to apply them for the Vietnamese language. In this paper, we mainly focus on testing the feasibility of applying UNL Vietnamese with these tools. First, we briefly introduce UNL and the system. Second, we introduce some tools that we have developed in research projects related to UNL. The final section presents experimental tools for the Vietnamese language. We especially want to mention translating a sentence from Vietnamese into UNL, and vice versa. From the experimental results, we show the orientation of further research that can apply UNL to developing automatic translation software for Vietnamese and ethnic minority dialects in Vietnam, such as Cham, Khmer, and Cơ-tu.

2. UNL Language and System

2.1. Introduction

The idea behind developing a multilingual translation system to use a pivot language that can cover the whole content of any natural language. In 1996, UNL was proposed by Dr. Hiroshi Uchida at the Senior Research Institute, United Nations University, Tokyo, Japan [1]. UNL has all the corresponding components of a natural language and includes word concept expressions called universal words (UWs), which link together to generate the UNL expression of a sentence. These links are called relations, which define the role of each word in a sentence. Showing the speaker's point of view is expressed through attributes.

- ***Universal Words***: Although UWs are mainly English, UWs include some words from other languages and semantic information to define the concept in the natural language. In this way, it helps to limit the inherent ambiguity of words in natural language. For example, the English word "state" will have two different words: `state (icl>country)` to represent the country, and `state (icl>region)` to denote a region of the country.

- ***Relations***: There are 56 relations in UNL, which are used to connect two UWs to build a semantic network of a UNL expression. These relations are the edges of the UNL graph or binary relations, which directly generate UNL expressions. For example, in a sentence like `"The boy eats potatoes in the kitchen"`, there is a verb `("eats")` and three arguments; two of them are instances of argumentative relations `("boy"` is the agent of the verb `"eats"`, whereas `"potatoes"` is the object) and one is a circumstantial relation (`"kitchen"` is the place where the action described in the sentence takes place).

- ***Attributes***: Attributes are used to describe the subjective information of a sentence. They clarify the speaker's point of view. The attributes include information about time or aspects of the event (number, polarity, modality, etc.). There are 87 attributes to clarify the semantics of the sentence.

In addition, the UNL system also uses a knowledge base so it can provide a semantic definition of the concept and confirm the extent of the relations to avoid definite ambiguity.

For example, the sentence `"John, who is the chairman of the company, has arranged a meeting at his residence"` is represented by the UNL language as follows:

```
{unl}
mod(chairman(icl>post):01.@present.@def,comp
any(icl>institution):02.@def)
aoj(chairman(icl>post):01.@present.@def,
John(icl>person):00)
agt(arrange(icl>do):03.@entry.@present.@comp
lete.@pred,John(icl>person):00)
pos(residence(icl>shelter):04,John(icl>perso
n):00)
obj(arrange(icl>do):03.@entry.@present.@comp
lete.@pred,meeting(icl>conference):05.@indef
)
plc(arrange(icl>do):03.@entry.@present.@comp
lete.@pred,residence(icl>shelter):04)
{/unl}
```

UNL expressions can also be represented in the form of UNL graphs. The UNL graph of the above-mentioned sentence is given in Fig. 2.

In Fig. 2, **agt** means the agent, **obj** is an object; **plc** is a place; **aoj** is an attributed object; and **mod** is a modifier. The detailed list of such relations can be found in the reference cited above. Also, the **icl** construct helps restrict the meaning of the word to a single meaning.

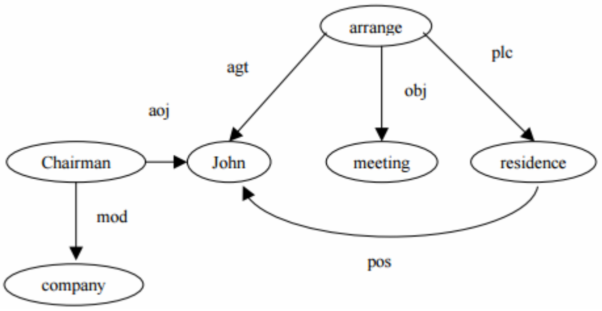


Fig. 2. Sentence expressed in terms of a graph.

2.2. UNL System

To integrate a language into the general UNL system, each natural language builds a language server on the Internet with the functions EnConverter and DeConverter. EnConverter and DeConverter are the core pieces of software in a UNL system. EnConverter is responsible for converting documents from the source natural language (NL) into UNL and DeConverter translates the text from UNL into the target NL. UNL servers can be connected to the Internet to carry out the conversions between natural languages and UNL expressions.

Anybody with access to the Internet can have EnConverter translate text in a source language into UNL expressions. The same UNL expressions can be translated into the target language via DeConverter. For the encoding and decoding process, we have to build a set of both encoding and decoding rules, and we need a bilingual UNL–NL dictionary.

The universal parser (UP) is a specialized version of EnConverter. It generates UNL expressions from annotated sentences by referring to the UW dictionary without using grammatical features. All UNL expressions are verified by the UNL verifier and are then stored in the format of a UNL document. DeConverter converts UNL expressions into natural language sentences. Both EnConverter and DeConverter perform their functions based on a set of grammar rules and a word dictionary of the target language. Consulting the UNL ontology and/or a co-occurrence dictionary in EnConverter or DeConverter is optional.

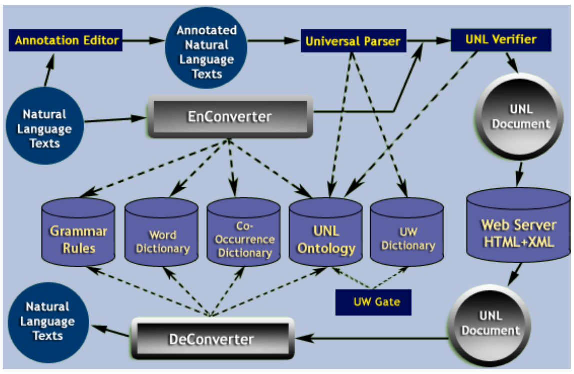


Fig. 5. Conversion mechanism of the UNL system.

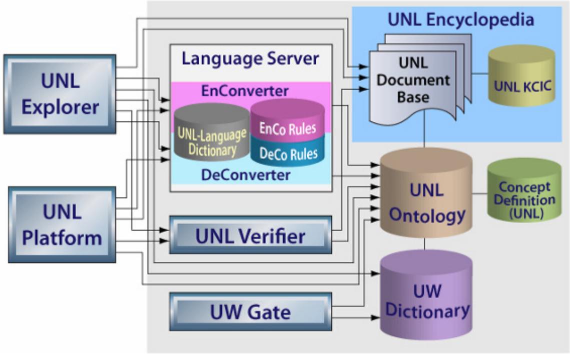


Fig. 5. UNL multilingual translation system.

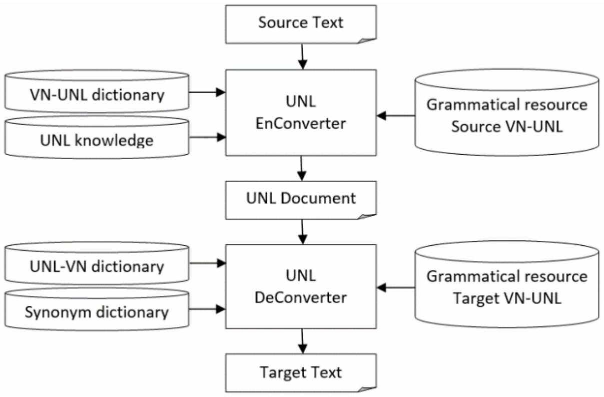


Fig. 3. Proposed model for the Vietnamese language.

Fig. 4 shows the structure of the UNL system and how it is connected to supporting tools and UNL-based applications. Highlighted parts are the components of the UNL system.

2.3. Support from UNL project

The UNL project has developed some tools to support adding a new language to the system. These tools are for semi-automatic language-independent analysis, but still need the support of a language specialist to resolve ambiguity. They will divide the input sentence, link correlative lexical items in the dictionary, and apply the transformation rules of grammar.

At the same time, we use UNL resources, such as UNL–NL dictionaries and UNL transformation, to process a source natural language into UNL documents, and process UNL documents into the target natural language. To use these tools, users must comply with the specified data format and format structures of grammar transformation rules. But there are still some languages that do not follow the rules, so these tools need adjusting.

3. Proposal for Vietnamese

Fig. 5 shows the mechanism how the conversion from a source text to the target text through the UNL system. We propose a system for the Vietnamese language based on the UNL system.

In this model, there is some important work to complete, such as the development dictionaries, grammatical resources, and tools for EnConverter and DeConverter.

The Vietnamese–UNL and UNL–Vietnamese dictionaries are built for use in EnConverter and DeConverter phases. They are important resources for integrating the Vietnamese language into the UNL system, and we developed them from 2010 to 2014 [3, 4].

The general syntax for the UNL–NL dictionaries is as follows:

```
[NLW]{ID} “UW”(ATTR,... ) < FLG,FRE, PRI >;
COMMENTS
```

with the following definitions.

- NLW: The lexical item of the natural language. Its format is decided by the dictionary builder. It can include:
- ID: The unique identifier (primary-key) of the entry.
- UW: A universal word in UNL, either simple ("book"), modified ("book.@pl") or complex ("aoj(new,book)"). This field can be empty if a word does not need a UW.
- ATTR: The list of features of the NLW.
- FLG: three-character language code according to ISO 639-3.
- FRE: The frequency of the NLW in natural texts. Used for natural language analysis. It can range from 0 (infrequent) to 255 (most frequent).
- PRI: The priority of the NLW. Used for natural language generation. It can range from 0 to 255.
- COMMENT: Any comment necessary to clarify the mapping between NL and UNL entries. It must end with the return code.

3.1. EnConverter

To perform the encoding process, we propose reusing an available tool called Interactive Analyzer (IAN).

IAN is software developed in the Web environment to perform encoding. Each natural language can be integrated and stored on the server. Therefore, data sources can be exploited anytime, anywhere, without being dependent on geographic distance. IAN analyzes the input with the help of T-Rules and dictionaries. IAN has eight tags: Welcome, NL Input, Dictionaries, N-Rules, T-Rules, D-Rules, IAN and Compare.

The tag “NL Input” allows users to provide documentation on the natural language. The tag “Dictionaries” allows users to provide NL–UNL dictionaries according to UNL specifications [9]. The tag “T-Rules” allows users to provide grammar conversion rules from the natural language into UNL. The tag “D-Rules” provides grammar orientation from the natural language into UNL. This tag is used as a control token and to improve the results of grammar transformation. The tag “IAN” shows the encoding result prescribed by UNL. The tag “Compare” is the tag to compare the results with other results.

Beside IAN, we can use EnCo for the encoding phase. EnCo is a tool used on a single machine [9]. Each input string is scanned from left to right and combined, taking entry in the word dictionary and it became the candidates. The candidates are sorted in order of priority. Selecting candidates is done by applying the rules of grammar. Parsing and semantics are done by applying the rules to choose words to build a syntax tree and a semantic network for the input. This process continues until all of the words in a sentence are put into and a complete semantic network for the input. The output of the entire process is a semantic network presented in UNL format.

3.2 DeConverter

To perform the decoding process, we propose reusing an available tool called EUGENE (from dEp-to-sUrface GENErator).

EUGENE is an application developed in the Web environment to create sentences in natural language from UNL expressions based on a UNL–NL dictionary and a set of transformation rules. EUGENE has seven tags: Welcome, NL Input, Dictionaries, T-Rules, D-Rules, EUGENE and Compare.

The tag “NL Input” allows users to provide documented action in the form of UNL expressions. The tag “Dictionaries” allows users to supply a UNL–NL dictionary according to the specifications of UNL [10]. The tag “T-Rules” provides grammar conversion and grammar orientation to convert from UNL into NL. This tag is used to control a token and improve the grammar transformation. The tag “EUGENE” shows that the result of encoding is a complete sentence in the natural language. The tag “Compare” is the tag to compare the results with others.

Besides EUGENE, we can use the DeCo tool, which is used on a single machine [10]. DeCo can deconvert UNL expressions into a native language, using Word Dictionary and Grammatical Rules. The DeCo tool works as follows. First, it transforms a UNL expression input into a directed graph with hyper-nodes called a node-net. The root node of a node-net is called the entry node and it represents the main predicate of the sentence. Second, it applies the decoding rules to nodes of the node-net and creates a word string in grammatical order of the target natural language. The decoding process ends when all the words to all the nodes are found, and a word string for the target language is completed.

4. Experiment for Vietnamese

With the help of tools like IAN and EUGENE from www.unlweb.net/wiki/Tools or www.undl.org/, along with the EnCo and DeCo tools, we can perform translation between any natural language and UNL. These include DeConverter, EnConverter, the word dictionary builder, the rules builder, and specifications or manuals of the tools.

Vietnamese is the isolating language; words in Vietnamese do not ingrain. The arrangement of words in a fixed order is the primary way to denote syntactic relationships in Vietnamese. For example, when we write: "Anh ta lại đến" it is a meaning different from "Lại đến anh ta"; or a combined word order such as "củ cải" is different from "cải củ", and "tình cảm" is different from "cảm tình".

To test for Vietnamese, we are going to build UNL–Vietnamese dictionaries and a rules set to provide to the tools. We encoded Vietnamese sentences into UNL and decoded UNL expressions into Vietnamese. Output was evaluated by language specialists for two criteria: to demonstrate the semantic meaning and the grammatical structure of the sentences. Next, we directly compared the quality of translations against the Google Translate tool based on the example sentences.

4.1 EnConverter with IAN

To provide the Vietnamese–UNL transformation grammar, we may either create a new file or upload an existing file. The transformation rules adhere to the following general formalism:

```
A:=B;
```

where the left side A is a condition statement, and the right side B is an action to be performed over A.

Example:

(%x):=(%x)(%y); the node %y is added to the right of the node %x

To encode a Vietnamese sentence “Tuấn đọc” using this tool, we need to provide:

- Input: “Tuấn đọc”
- After analyzing the above sentence, we have the following result from the IAN tool: [Tuấn][][đọc]
- The Vietnamese–UNL dictionary for the above sentence will be:

```
[tuấn]{}"tuan(icl>name,iof>person,com>
male)"(LEX=N,POS=PPN,SNGT) <Vie,0,0>;
[đọc]{}"read(icl>see>do,agt>person,obj>infor
mation)"(LEX=V,POS=VER,att=@present)<vie,0,0
>;
```

- Use rules to encode the above sentence:

```
R1.(%a,BLK):=;
R2.(N,NT,%a)(V,Vt,att=@present,%b):=(  agt(%
b,+att=@entry;%a));
```

- Output: a UNL expression as created correspondence:

```
### IAN Console ###
1.Pattern:  [Tuấn] [] [đọc]
[S:1]
{org}
Tuấn đọc
{/org}
{unl}
agt(read(icl>see>do,agt>person,obj>informato
n):03.
@present.@entry,tuan(icl>name,iof>person,com
>male):01)
{/unl}
[/S]
```

4.2. EnConverter with EnCo

With the EnCo tool, syntactic and semantic analyses are carried out by applying the enconversion rules, and the relevant operations are conducted according the rule type. Syntactic analysis is carried out by applying rules of types that are either “+”, “-”, “<” or “>”; semantic analysis is triggered by the indication of relations in the “relation.”

The enconversion rule has the following syntax:

```
<TYPE>"{" [<COND1>] ":" [<ACTION1>] ":" [<REL1>]
":" "}""{" [<COND2>] ":"
[<ACTION2>]":"[<REL2>]":""}"("<PRIORITY> ");"
```

in which

<COND1> and <COND2> show condition 1 and 2, which contain the lexical attributes and semantic left and right analysis windows.

<ACTION 1> and <ACTION 2> show actions that are only performed if the corresponding conditions are true.

<REL1> and <REL2> show possible relations between the two analysis windows.

<PRIORITY> describes the interpreted order of the rules.

For example:

```
>{N: null: aoj}{ADJ:+R:null};
```

This is a right modification rule. The results removed the left button of the node-list and by adding the attribute *R* to the attributes of the right analyze window for further processing. The action field of the left analysis window contains *null*. It is not necessary to do anything in the left analysis window. The relation field of the left analysis window contains *aoj*, and the relation field of the right analysis window contains *null*. As a result, it solves the relation of *aoj* between the left analysis window and the right analysis window.

To encode a Vietnamese sentence “Tuấn đọc” using this tool, we need to provide:

- Input: "Tuấn đọc"
- The Vietnamese–UNL dictionary for above sentence:

```
[tuấn]{}"tuan(icl>name,iof>person,com>
male)" (N,PN)<Vie,0 ,0>;
[đọc{}"read(icl>see>do,agt>person,obj>
information)"(V,VT,SNGT)<vie,0,0>;
```

- Use rules to encode the above sentence:

```
R1.R{}{^mor,^STAIL:mor};
R2.:{}{^mor,STAIL:mor};
R3.R{}{mor};
R4.+{^blk:blk}{BLK};
R5.:{}{mor,N,SUBJ,^EONP:EONP,hnp};
R6.DR{N,PN}{BLK};
R7.:{}{mor,V,^EN,^&@present.@entry,
&@present.@entry};
```

```
R8.>{hnp,^prep:&@topic:agt}{V,EOR,^>agt,^sub
j:subj};
R9.:{^EOR:EOR}{STAIL};
```

- Output: UNL expression as created correspondence:

```
[S:1]
{org}
tuấn đọc
{/org}
{unl}
[W]
:00
[/W]
{/unl}
{unl}
agt(read(icl>see>do,agt>person,obj>informati
on):0A.
@present.@entry,tuan(icl>name,iof>person,com
>male):03.@topic)
{/unl}
{/S}
;;Time 0.0   Sec
;;Done!
```

4.3 DeConverter with EUGENE

The rules are built in the same way as for IAN. For example:

(%x,N,@def):=(NS(%x,@def;%y,[chiếc],+LEX=D,+POS =ART));

Remove the attribute @def, create new relations "NS" with two UWs "%x" và "chiếc" where the attribute is (LEX = D, + POS = ART).

To decode a UNL expression using this tool, we need to provide:

- Input:

```
{unl}
agt(read(icl>see>do,agt>person,obj>
information).@entry.@past,tuan(icl>name,iof>
person,com>male))
{/unl}
```

- The Vietnamese–UNL dictionary for the above sentence:

```
[tuấn]{}"tuan(icl>name,iof>person,com>
male)"(LEX=N,POS=PPN,NUM=SNGT)<vie,0,0>;
[đọc]{}"read(icl>see>do,agt>person,obj>infor
mation)"(LEX=V,POS=VER)<vie,00>;
```

- Use rules to encode the above sentence:

```
R1.agt(%x,V,@past;%y,N):=VS(%x;%y,+CAS=NOM);
R2.(%x,V,@past):=PS(%x,@past;%y,[đã],
+LEX=D, +POS=ART);
R3.VS(%x;%y):=(%y,+>BLK)(%x,+>BLK);
R4.PS(%x;%y):=(%y,+>BLK)(%x,+>BLK);
R5.(%x,>BLK)(%y,^BLK,^PUT,^STAIL):=(%x,>BLK)
("",+BLK)(%y);
```

- Output : Vietnamese is created:

```
### please select the sentence to run from
the list above ###
[S:1]
{org}
tuan read
{/org}
{eng}
tuấn đã đọc
{/eng}
{unl}
agt(read(icl>see>do,agt>person,obj>informati
on).@entry  .@past,tuan( icl>name,iof>person
,com>male))
{/unl}
[/S]
```

4.4. DeConverter with DeCo

The rules are built in the same way as for EnCo. For example:

"HPRON,SUBJ:subj:agt"{V,^IRG,^pred:pred}; insert object of relations "AGT" into the node list.

To decode a UNL expression using this tool, we need to provide:

- Input:

```
[S:1]
{org}
I wrote
{/org}
{unl}
agt(write(icl>communicate>do,agt>person,obj>
information,ins>thing,rec>person) .@entry.@p
ast,I(icl>person))
{/unl}
[/S]
```

- The Vietnamese–UNL dictionary for the above sentence:

```
[tôi]{}"I(icl>person)"(PRON,HPRON,SUBJ)<vie,
0,0>;
[viết]{}"write(icl>communicate>do,agt>
person,obj>information,ins>thing,rec>
person)"(V,VT,SNGT)<vie,0,0>;
```

- Use rules to encode the above sentence:

```
R1.R{}{};
R2.:"HPRON,SUBJ:subj:agt"{V,@past,ED, ^pred:
pred};
R3.:"HPRON,SUBJ:subj:agt"{V,@past,ED, ^pred:
pred};
```

- Output: Vietnamese is created:

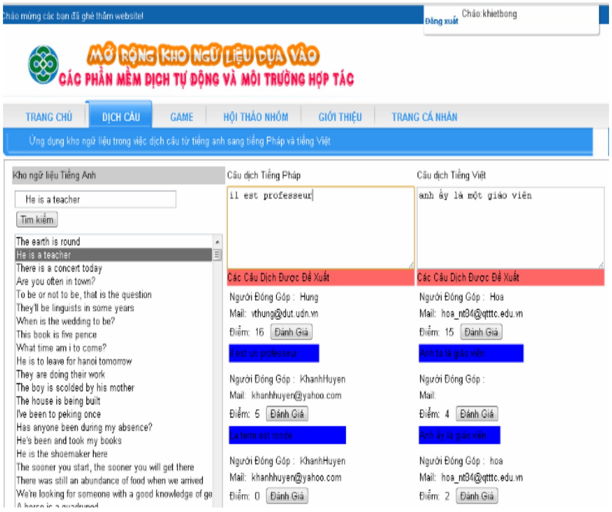
```
[S:1]
============WORD LIST :02===========
01 00 00:<agt
[tôi]{} “I(icl>person)”(HPRON,PRON,SUB)
<Vie,0,0>;
02 00 00:@entry,@past,>agt
[đã viết]{}“ write( icl>communicate > do,
agt > person, obj> information,ins> thing,
rec > person)” (V,VTSNGT) <Vie,0,0>;
============UNL===========
write(icl>communicate>do,agt>person,obj>info
rmation, ins>thing, rec
>person)(@entry,@past,>agt)
===================================
tôi đã viết
;;Time 0.0   Sec
;;Done!
```

5. Performance Evaluation

EnCo and Deco have interfaces as well as the encoding results, which are unfriendly to users and software used on a single computer; it is difficult to share data with other users. IAN and EUGENE are built on web platforms and with a friendly interface. Special resources like dictionaries and rules can be shared with other users in the online community. However, with IAN and EUGENE, we need semantic disambiguation, and then, the system filters the candidates using the optional rules set for grammar, called D-rules.

Table 1. The change rate between two machine translations.

Translation method	Change rate of sentences	
	No change	Change
Vietnamese → UNL → Vietnamese	746 (74.6%)	254 (25.4%)
Vietnamese →English →Vietnamese	653 (65.3)%	347 (34.7%)



Chào mừng các bạn đã ghé thăm website!

Đăng xuất Chào:khietbong

MỞ RỘNG KHO NGỮ LIỆU DỰA VÀO CÁC PHẦN MỀM DỊCH TỰ ĐỘNG VÀ MÔI TRƯỜNG HỢP TÁC

TRANG CHỦ | DỊCH CÂU | GAME | HỘI THẢO NHÓM | GIỚI THIỆU | TRANG CÁ NHÂN

Ứng dụng kho ngữ liệu trong việc dịch câu từ tiếng anh sang tiếng Pháp và tiếng Việt

Kho ngữ liệu Tiếng Anh

He is a teacher

Tìm kiếm

The earth is round
He is a teacher
There is a concert today
Are you often in town?
To be or not to be, that is the question
They'll be linguists in some years
When is the wedding to be?
This book is five pence
What time am i to come?
He is to leave for hanoi tomorrow
They are doing their work
The boy is scolded by his mother
The house is being built
I've been to peking once
Has anyone been during my absence?
He's been and took my books
He is the shoemaker here
The sooner you start, the sooner you will get there
There was still an abundance of food when we arrived
We're looking for someone with a good knowledge of ge
A horse is a quadruped

Câu dịch Tiếng Pháp

il est professeur

Các Câu Dịch Được Đề Xuất

Người Đóng Góp : Hung
Mail: vthung@dut.udn.vn
Điểm: 16 Đánh Giá
il est un professeur

Người Đóng Góp : KhanhHuyen
Mail: khanhhuyen@yahoo.com
Điểm: 5 Đánh Giá
La terre est ronde

Người Đóng Góp : KhanhHuyen
Mail: khanhhuyen@yahoo.com
Điểm: 0 Đánh Giá

Câu dịch Tiếng Việt

anh ấy là một giáo viên

Các Câu Dịch Được Đề Xuất

Người Đóng Góp : Hoa
Mail: hoa_nt84@qtttc.edu.vn
Điểm: 15 Đánh Giá
Anh ta là giáo viên

Người Đóng Góp :
Mail:
Điểm: 4 Đánh Giá
Anh ấy là giáo viên

Người Đóng Góp : hoa
Mail: hoa_nt84@qtttc.edu.vn
Điểm: 2 Đánh Giá

Fig. 6. User interface to evaluate sentences

We tested the tools with some Vietnamese sentences (for example, encoding a Vietnamese sentence into UNL, and UNL into Vietnamese) with satisfactory output for translation quality. We will continue to expand the UNL–Vietnamese dictionary and share it with the online community.

Moreover, to evaluate the translation results, we translated 1000 sentences from Vietnamese in two ways.

The first was translation by Google Translate from Vietnamese to English and then from English to Vietnamese. This means that we used English as the pivot language with statistical machine translation systems.

The second was translation by a UNL system from Vietnamese to UNL through the IAN tool, and then from UNL to Vietnamese via the EUGENE tool. The rate of sentence change when comparing source and target sentences is as follows:

With the support of language experts, we will compare the results of the two systems based on two characteristics: semantics and syntax.

To support the experts in assessing the quality of the translated sentences, we developed software that displays the original sentence and its translation from the automation software, and language experts then comment on the translation quality of each sentence.

The software interface is in Fig. 6.

Table 1 shows that the rate of Vietnamese sentences changed when using UNL is lower than when using Google Translate and English as an intermediate language. This is better than natural language (for example, English).

The reason for this difference is that the use of natural language as an intermediary language is an unbreakable barrier in the knowledge representation of a language; UNL performs with all the knowledge of the natural language without encountering problems, such as multi-semantics of a word, semantic ambiguity, grammar, and context dependence.

6. Conclusion

Through experiments, we find that conversion tools from natural language to UNL and vice versa are effective. The quality is acceptable. However, to use these tools, users must comply with the specified data format. But there still are some languages that do not follow the rules, so these tools need adjusting.

In this paper, we introduced UNL application tools for automatic Vietnamese translation. We also demonstrated how a sentence can be converted into a UNL expression, and vice versa, by taking a sentence as an example. Although we have already worked with examples, our implementation must cover all dialects of Vietnamese, but that is the start of a long journey, and we hope that we can reach that destination.

We see that the UNL system is a promising way to develop a multilingual translation system for natural language. It allows for the precise expression of any specific concept of a particular language, without ambiguities, which is understandable for computers and humans using other languages. This system allows any natural language to participate, and we only developed two encode and decode modules.

We have built a large enough Vietnamese–UNL dictionary. In the future, we will install an automatic translation Vietnamese server with two functions: EnConverter and DeConverter. In the subsequent stages, we will continue to install the new server for the dialects of ethnic minorities in Vietnam, such as Chăm, Khmer, and Cơ-tu, in order to have a complete multilingual automatic translation system for the languages used in Vietnam.

References

[1] H. Uchida, "UNL: Universal Networking Language – An Electronic Language for Communication, Understanding, and Collaboration", *UNU/IAS/UNL Center*, Tokyo, Japan, 1996. Article (CrossRef Link)

[2] S. Tripathi and J. K. Sarkhel, "Approaches to machine translation", *Annals of Library and Information Studies*, vol. 57, 2010, pages 388-393. Article (CrossRef Link)

[3] P.T.L.Thuyền, V.T.Hung, "Expansion of UNL – Vietnamese dictionary on UNL explorer", *Journal of Science and technology*, The University of Danang. N.11(84), 2014. Article Article (CrossRef Link)

[4] V. T. Hung and G. Fafiotte, "UVDict – a machine translation dictionary for Vietnamese language in UNL system", *Proceeding CISIS 2011*, Korean Bible University (KBU), Seoul, Korea, 2011, Pages: 1020-1028. Article (CrossRef Link)

[5] V. T. Hung, "Reuse of Free Online MT Engines to Develop a Meta-system of Multilingual Machine Translation", *EsTAL 2004*, 2004. Article (CrossRef Link)

[6] V. T. Hùng, "Phương pháp và công cụ đánh giá tự động các hệ thống dịch tự động trên mạng", *Tạp chí Khoa học và Công nghệ Đại học Đà Nẵng*, số 1 (18), 2007, tr. 37-42. Article (CrossRef Link)

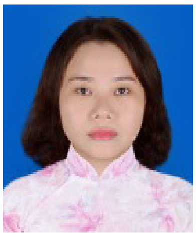
[7] P. Kumar and R.K. Sharma, "Punjabi to UNL EnConversion System", *Springer: Sadhna, Academy Proceedings in Engineering Sciences*, volume 37:(2), pp. 299-318, 2012. Article (CrossRef Link)

[8] P. Kumar and R.K. Sharma, "Punjabi DeConverter for generating Punjabi from Universal Networking Language", *Journal of Zhejiang University - Science C 14(3)*: 179-196 (2013*)*. Article (CrossRef Link)

[9] UNL centre, "EnConverter Specifications. Version 3.3", 2002. Article (CrossRef Link)

[10] UNL centre, "DeConverter Specifications. Version 2.7", 2002. Article (CrossRef Link)

Vo Trung Hung is a Professor of Computer Science at the University of Danang (UD), Vietnam. He is the director of Science and Technology Department of the UD. He received his Ph.D. in Computer Science from Institut National Polytechnique de Grenoble (INPG), France, in 2004. His current research interests include natural language processing, automatic translation, and multilingual softare.



Phan Thi Le Thuyen graduated B.A. from College of Science-Hanoi Na-ional University and M.A. degree in Computer Science from the Univer-ity of Danang. She is now a Ph.D. student in Computer Science of the University of Danang, Vietnam. Her thesis is on the development of multilingual tran-lation system.