The Treatment of Lexical Ambiguity in Machine Translation

Asst. Prof. Dr. Tawfiq Aziz Abdullah ∗
Lect. Safwan Idrees Thanoon ∗

Abstract

This research sheds light on one of the most difficult problems especially in designing a translation programme which deals with ambiguity. The programme can be built in a form of categories of words that cannot be adjacent in a sentence.

From a semantic point of view, a word certainly has different meanings depending on situations and contexts. It is known that a computer as a machine cannot think as a human being. Thus, our aim is to find a best way of analysing a sentence in a mathematical way, i.e. to find out a programme that uses data bases, calculations, modules that can be linked to each other for specifying the lexical ambiguity.

1. **Introduction:**

Machine translation (MT) "is the area of information technology and applied linguistics dealing with the translation of human languages such as English and Arabic" (Shaalan et al., 2004:2 ; Ramanathan,A., undated:3).

Ambiguity exists in all languages. It certainly poses problems in the conception of applying automatic translation and it is the principal factor that limits their performances.

It is very important to think that a computer is not a 100 percent intelligent machine. Therefore, we must treat language in a logical

Dept. of French/ College of Arts / University of Mosul .
Dept. of Translation / College of Arts / University of Mosul .

manner, getting use of the capacity of computer in doing calculations very fast.

We must transpose the traditional methods of translation which appeal to the intelligence of a translator. Certain things that are evident to us need a lot of work. We have to avoid ,therefore, many models before we pass writing the codes because it seems that a simple problem requires certain lines of programming.

It is better as it seems to begin our model and programming with the simplest aspect of analysing the language, i.e. its regularities. After that we have to take into consideration the problems and to find a solution even if we are far away from a traditional grammar. This aspect is important for the one who starts to realize this designing of automatic translation.

After this short introduction, it is time to study the syntactic and semantic ambiguities and the eventual methods of programming which are going to be dealt with. Perhaps, there are many methods. Our method ,however, is trying to solve the problem.

In this study we shall use the (Golden Al Wafi, version 1.12) machine translator to verify our results.

2. <u>Syntactic ambiguity</u>:

A word which has a particular form can be interpreted in different manners. This is the case ,for example, with the word 'sprigs' which can be analysed as a plural noun or a verb in the simple present (third person singular). The programme will carry a list of categories which cannot be found in any case beside each other. Thus, it avoids certain ambiguities. Here is one description among many:

-        **Article + Verb**: In English, this combination is impossible but nouns in the plural could have the same form rather than a verb in the present simple with a third person singular. We can give another example showing this kind of ambiguity:

1. He likes the parks of Baghdad.

The word park can be analysed like the following: The verb 'to park' with the simple present or the noun 'parks' in the plural. However, the Golden Al Wafi avoids the possibility of article + verb and only the solution of the word in the plural will be caught:

IL aime les parkings de Bagdad. (French)   يحب متنزهات بغداد

- Possessive + Noun:

2. Have you read his books?

   Aves- vous lu ses livres?   (French)   هل قرات كتبه؟

The word 'books' can be a noun or a verb and the possessive pronoun will remove the ambiguity.

- Personal pronoun + Noun:

3. He parks the car.

      IL gare la voiture.   (French)   يوقف السياره

It is impossible to have a personal pronoun besides a noun like English and French, and the noun will remove the ambiguity on 'parks' which is analysed as a verb in the present simple.

 - ∅ + **Possessive**: It is possible to remove the grammatical ambiguity in the following sentence:

   4.Who did that? Her.

In this example, the word 'her' is assimilated to a sentence (presence of a an interrogative mark before and full stop just after). For more about the features of the noun phrase see (Kurland's,D, 2003:Internet).

The fact is that 'her' alone will allow the machine to avoid the possibility of possessive adjective.

The above argument shows clearly that some machine translators can remove the ambiguity in the previous sentences because it distinguishes between verbs and nouns. There is a list of meanings of a given word. This means that it can easily detect whether a certain word in a sentence is a noun or verb, i.e. there is no big problem.

Concerning the structural ambiguity in the sentence:

5. Flying planes can be dangerous.

the word 'flying' can be interpreted either as an adjective functioning as premodifier of the noun 'planes' or as a verbal noun (Lyons, 1968:249).

Usually two renderings can be provided for this sentence by any translator depending on the context of situation which removes this ambiguity

الطائرات المحلقة قد تكون خطرة

تحليق الطائرات قد يكون خطراً

Les avions volants euvent etre dangereux. (French)

Thus, if sentence (5) to be:

6. Flying planes <u>is</u> dangerous.

   Les avions volants sont dagereux (French)

 the translation of volants will be 'تحليق', i.e. into a verbal noun whereas if sentence (5) is :

7.Flying planes <u>are</u> dangerous.

   Les avions volants sont dangereux. (French)

then 'المحلقة' is more appropriate because 'are' in (7) belongs to 'planes' while 'is' in (6) belongs to 'flying'.

The problem in this argument is that the machine translator gives only one choice 'الطائرات المحلقة' for sentences (5), (6), and (7). This means that the machine itself creates ambiguity even in the sentences the ambiguity of which is already resolved. It cannot deal with this structural ambiguity properly because the verbal noun 'flying' combines the features of both nouns and verbs (see Scheurweghs,1959:175ff).

1. <u>Semantic ambiguity</u>:

       Semantic ambiguity is considered more complex than the syntactic one because only the intelligence of human being is capable of translating such ambiguity. For example, in the sentence:

8. He went to the bank.

 IL est alleʹ a la banque / IL est alleʹ a la bord   (French)

    It is not clear whether the word 'bank' refers ,for example, to the 'mound of sand' or to the 'financial institution' (see Palmer, 1980:106; Shaalan et al., 2004:4). In this sentence only the context of situation helps in determining the choice. The machine translates this sentence into:

ذهب الى المصرف

However, the following sentences:

9. He went to the bank and sat under a tree.

10. He went to the bank to save money.

can be translated easily by a translator ,respectively, into:

ذهب الى ضفة النهر وجلس تحت شجرة

ذهب الى المصرف لتوفير المال

but the problem is that when inserting (9) and (10) into the

computer, only the word 'مصرف' appears for both cases:

ذهب الى المصرف وجلس تحت شجرة

ذهب الى المصرف لتوفير المال

This means that the computer could not distinguish between the two meanings of (9) and (10) even if a context exists.

What is needed here is to make the machine capable of dealing with the given context of a certain text. In this study we find a best logical model for the computer to do less possible faults. This can be made possible by creating a database, and we have to develop what is called 'motor of reading' which has the ability to exploit this base. The database must be done in order to be adapted with this motor of reading. We make groups of possible key words and the machine will automatically choose the suitable word.

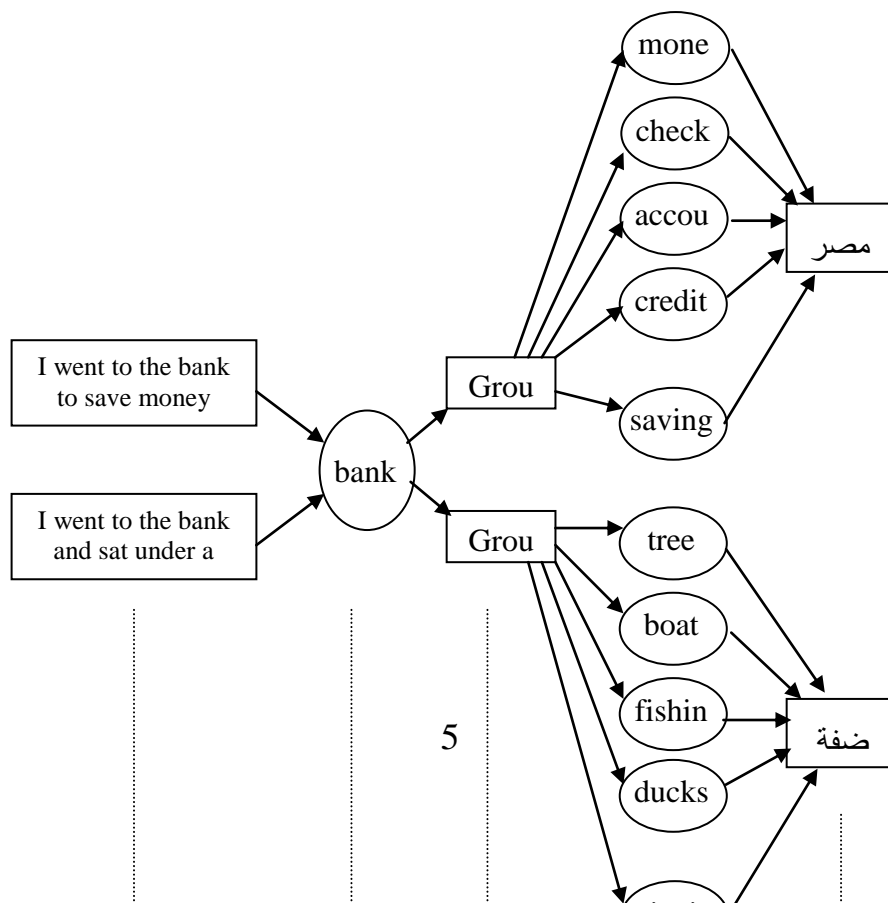Our following module (Figure 1) explains sentences (9) and (10):

5

Figure 1 (suggested module)

Another example would be the word 'plant' which can denote 'vegetation growing from the ground' and can denote 'factory' (McNaughton, 1082:231). In general, we can apply this to all the possible ambiguous words of a language in the same way, i.e. by giving groups of key words. Our following algorithm and flowchart explain the suggested programme:

Step 1: Enter the text.
Step 2: Search for ambiguous word.
Step 3: If found, search in the key words groups for a word. If not exit.
Step 4: If one of the key words in the first group is included in the entered text, translate the ambiguous word depending on the meaning given in the first group. Otherwise, translate the ambiguous word in other groups and so. Go to step 6.
Step 5: If the text has not any of the key words in all groups then exit.
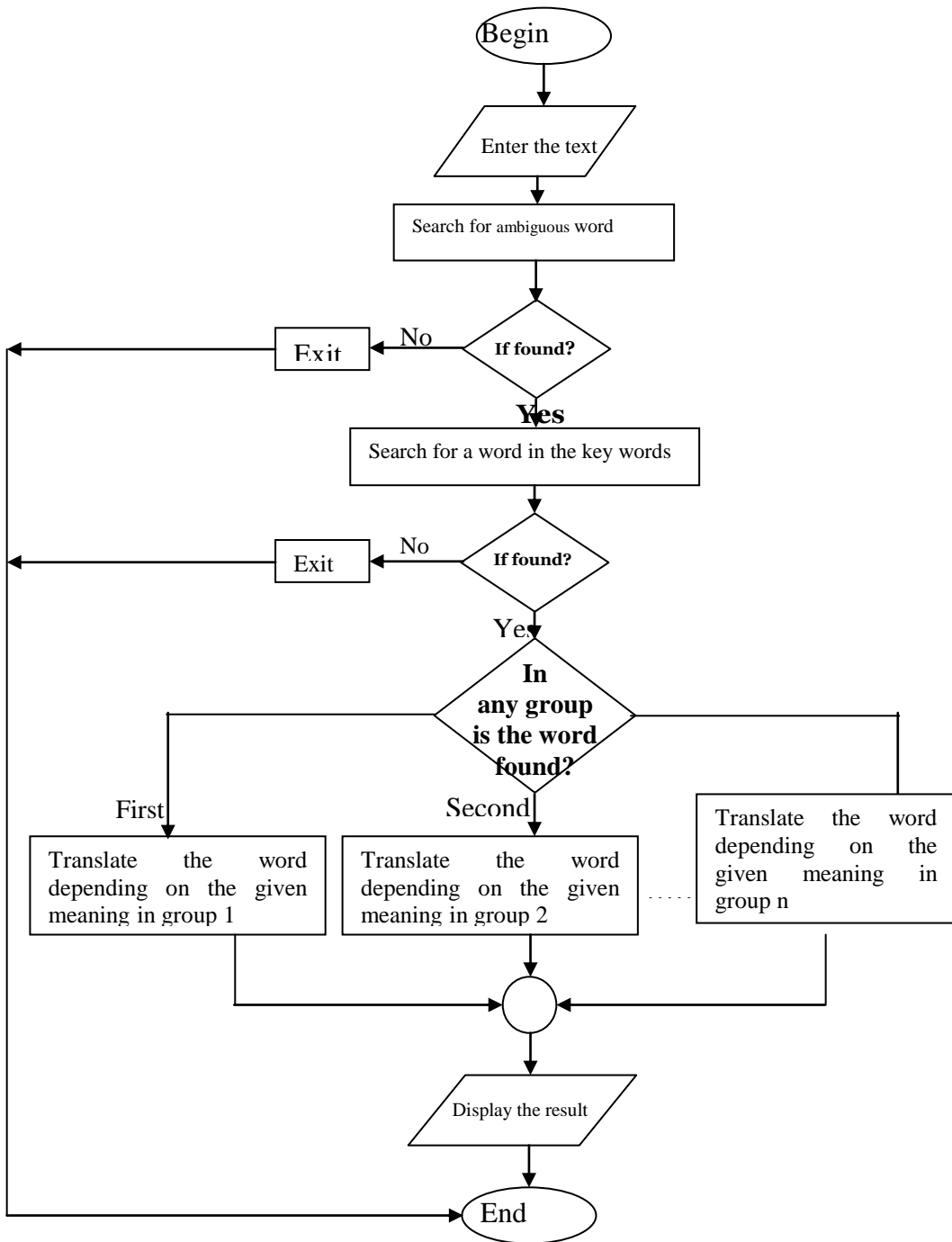Step 6: Display the result on the screen.

Begin

Enter the text

Search for ambiguous word

If found? — No → Exit

Yes

Search for a word in the key words

If found? — No → Exit

Yes

In any group is the word found?

First

Translate the word depending on the given meaning in group 1

Second

Translate the word depending on the given meaning in group 2

Translate the word depending on the given meaning in group n

Display the result

End

Figure (2)

This programme is done by using MATLAB version 7.

2.   Problems of translating the verb phrase in machines:

Although sometimes a context exists, it is difficult for the machine to distinguish the meaning of a certain word. The following examples of the verb 'hold' illustrates the different meanings of this verb:

11. He holds some books.
12. He holds silence.
13. He holds a theory.

IL porte quelques livres (French)
IL tient la silence.
IL maintient une theorie.

The machine translates them ,respectively, into:

يحمل بعض الكتب

يحمل صمتاً

يحمل نظرية

However, the word 'hold' has not the same meaning in the three sentences and the translation of it should not be the same. More appropriate translations of (12) and (13) would be يلزم الصمت and يؤمن بنظرية (see Ba'lbakki, 1990:430).

Another example is the adjective 'loaded':

14. The lorry is loaded.
15. The merchant is loaded.
16. The gun is loaded.

The machine gives the following translations:

ان الشاحنة محملة

ان التاجر محمل

ان البندقية محملة

Again the word 'loaded' has different meanings in the different sentences and ,therefore, (15) and (16) can better be translated into البندقية محشوة and التاجر غني . In French, (14), (15), and (16) can be rendered ,respectively, into (Le camion est charge′), (Le commercant est riche), and (Le fusil est charge′).

Usually, the meaning of the verb phrase is determined by its relation with the words but examples like these need to be treated. The solution is difficult but not impossible which is by creating a list of words or verbs which can be the database. In this case, every time the computer finds the verb or the word, it considers if it is in the database and it will compare it with the elements of the sentence. We can imagine that the mass of data is necessary for all the English words or at least words that are mostly used. Of course the machine succeeded in translating many verbs and words.

There is a more complicated kind of ambiguity in translation because it needs a modelisation which must stimulate a treatment of intelligent language. Let us take ,for example, the adjective 'cold':

17. She was cold because she walked in the snow.

اصابها البرد لانها سارت في الثلج

Elle avait froid parce qu'elle a marche dans la neige. (French)

18. She was cold with her husband.

كانت تتعامل ببرود مع زوجها

Elle etait froide avec son marie . (French)

Certainly, the expression (to be cold) in the above sentences is translated differently. This is due effectively to the context in which these expressions are found. This leads us to say that we have to find a precise word for this ambiguous word. In such case, we have to look for the context. To simplify a situation which depends normally on actors, acts, ways,…is not an easy way to remove the kind of ambiguity in machines but it is not impossible.

5. <u>conclusion:</u>

Removing ambiguities is the most important factor for a programme of translation. It is important to start a work with a theory before starting any programme.

We think it is better to invent adapted rules even if they are absurd but the goal is to have a result.

The only chance to remove ambiguity depends on the power of calculations which allow to calculate a great number of possibilities and to have an effective modelisation of ambiguities. However, increasing the number of key words in any suggested dictionary improves the execution of a programme. This ,of course, requires a longer storage memory.

References

Ba'lbakki,M, (1990). Al Mawrid (English-Arabic Dictionary). Dar Al Elm Lilmalaeen: Beirut.

Kurlands,d. (2003) How the Language Really Works: The Fundamentals of Critical Reading and Effective Writing.
Internet: httpwww.criticalreading.com/nouun-phrase.htm

Lyons,J. (1968) Introduction to Theoretical Linguistics. Cambridge: Cambridge University Press.

McNaughton,R. (1982) Elementary Computability, Formal Languages, and Automata. Prentice-Hall, INC., New Jersey.

Palmer,F (1981) Semantics. Cambridge: Cambridge University Press.

Ramanathan,A (undated) "Statistical Machine Translation". Indian Institute of Technology, Bombay Mombay.1-33.
Internet: http://www.cse.iitb.ac.in/~anand/statmt.pdf

Shallan,K. (2004) "Machine Translation of English Noun Phrases into Arabic". International Journal of Computer Processing of Oriental Languages. Vol. 17, No. 2. 1-14.

Scheurweghs,G. (1959) Present-Day English Syntax, Longman Group Limited: London.

معالجة غموض المفردات في الترجمة الآلية

أ.م.د.توفيق عزيز عبدالله

م.صفوان أدريس ذنون

مستخلص

يسلط هذا البحث الضوء على واحدة من اصعب المشكلات فيما يخص تصميم برنامج ترجمة لمعالجة الغموض. ويبنى البرنامج على شكل فئات من الكلمات التي لا يمكن ان تكون متجاورة في الجملة.

تمتلك اية مفردة من الناحية الدلالية معاني مختلفة بالاعتماد على السياق الذي تقع فيه. ومن المعروف بان الحاسوب ليس لديه القدرة على التفكير كالبشر باعتباره آلة ولهذا فان هدفنا هو ايجاد افضل وسيلة لتحليل الجملة بطريقة رياضية أي بايجاد برنامج يستخدم قواعد بيانات وحسابات ووحدات برمجية بحيث يمكن ربطها مع بعضها البعض لمعرفة الغموض.