

A Lexical-Functional Model for Machine Translation of English Zero-place Predicators into Persian

Abbas Ali Ahangar¹, Nader Jahangiri² & Fahimeh Mohammadpour¹

¹ University of Sistan and Baluchestan, Zahedan, Iran

² Ferdowsi University of Mashhad, Mashhad, Iran

Correspondence: Fahimeh Mohammadpour, University of Sistan and Baluchestan, Zahedan, Iran. Tel: 98-915-846-9037. E-mail: fahime_mohammadpour@yahoo.com

Received: March 19, 2012

Accepted: April 17, 2012

Published: June 1, 2012

doi:10.5539/ijel.v2n3p2

URL: <http://dx.doi.org/10.5539/ijel.v2n3p2>

Abstract

Comparative analysis of English and Persian sentence structures reveals that though zero-place predicators are one of predicator types in English, they do not exist in Persian. So, generating a natural translation of them especially in machine translation systems is essential. This study aimed to show that utilizing Lexical-Functional Grammar for a machine translation system which is designed for translation of some English zero-place predicators is able to produce a natural translation of them into Persian. To accomplish this purpose, after determination of part of speech of words of the sentences in C-structure and grammatical function of noun phrases present in the source language (i.e. English) sentence in F-structure, the suitable equivalence was selected for each word and finally the sentence was generated in target language (i.e. Persian). The findings of the present research suggest that English zero-place predicators translated into Persian by a lexical-functional machine are more natural and also they are based on Persian word order.

Keywords: machine translation, C-structure, F-structure, A-structure, source language, target language

1. Introduction

Translation is an old story for human being. But ever-increasing information flow has changed the type of human being's need to translation. Nowadays, no one expects when s/he is reading an internet page and does not understand some parts or whole of it; a human translator sits down next to her/ him and translates the text. This big change in every one's attitude and high speed of information flow in the modern world has led to thinking of creating a system which is able to translate the given text automatically and independently. This system has been called machine translation.

The study of the short history of machine translation systems shows that these systems have always been designed based on a special linguistic theory. In other words, linguistics can be regarded as an inseparable part of each machine translation system.

Younesi Far (1994) designed and built a machine translation system based on Augmented Transition Networks / ATN. The given system which used the sentence model of Wishon could produce a word for word translation of English sentences into Persian. Another machine translation system for translation from English into Persian is a rule-based machine translation based on HPSG. In this system, meaning is shown by MRS semantic structure (Niknejad, 2008). Saedi (2008) introduced a hybrid machine translation for translation of simple English sentences into Persian. He (ibid) believed that disambiguation and transfer are the most important parts of this machine translation. Faroughi (2007) conducted a machine translation system based on lexical-functional grammar. The given system was designed for translating English sentences into Persian in general with a specific focus on translation of noun phrases.

So, the present study aims at showing that utilizing Lexical-Functional Grammar for a machine translation system which is designed for the translation of some English zero-place predicators is able to produce a natural translation of them into Persian.

In this regard, in what follows, in the first part, machine translation system and Lexical-Functional Grammar will be described. Then, the method of the research is presented. Finally, the article ends with the results and discussion.

1.1 Machine Translation / MT Systems

MT is one of sub-branches of computational linguistics. In the later half of the twentieth century, computational linguistics, as a new branch of applied linguistics, arose for construction of computer programs to process words and texts in natural language (Bolshakov & Gelbukh, 2004).

Another definition for computational linguistics has been given as “the analysis of written texts and spoken discourse, the translation of text and speech from one language into another, the use of *human* (not computer) languages for communication between computers and people, and the modeling and testing of linguistic theories” (Fromkin, Rodman, & Hyams, 2007: 412).

Hutchins (2003: 501) defines machine translation as “computational systems responsible for the production of translations with or without human assistance.”

Generally, three approaches were used in MT systems before 1990s:

1). Direct Approach: In this approach, “systems were designed in all details especially for one pair of languages, i.e. in most cases, for Russian as SL and English as TL. The basic assumption was that the vocabulary and syntax of texts should be analyzed no more than necessary for the resolution of ambiguities, the correct identification of appropriate TL expressions and the specification of TL word order.” (Hutchins, 1979: 31)

Aasi (2004: 34) believes that since, in this approach, translation process is done through replacement of the equivalent words, MT acts like a mechanical dictionary. MT systems which employ the direct approach are grouped as the first generation of MT systems.

The Georgetown University system, demonstrated in 1951, was typical of the ‘direct’ approach, because it “illustrates well the complexities and the ultimately insuperable problems of the ‘direct’ approach.” (Hutchins, 1979: 31-32)

2). Interlingua Approach: This approach “assumes that it is possible to convert SL texts into semantico-syntactic representations which are common to more than one language (but not necessarily ‘universal’ in any case). From such interlingual representations texts are generated into other languages. Translation is thus in two stages: from SL to the interlingua and from the interlingua to the TL.” (Hutchins, 2003: 503) Hesabi (2006: 43) names this representation as “representation schema which is to some extent independent from SL and TL.”

Hutchins (1979) supposes that Warren Weaver was the first one who mentioned the attractiveness of an interlingua approach to MT in his famous memorandum. “But it was not until the 1960’s when theoretical linguistics has turned to problems of language universals that MT researchers had any clear ideas of how interlinguals could be constructed.” (ibid: 33)

Though Hutchins (1979) sees Interlingua approach positively, Wilks (2009:122) criticizes this approach as: “an interlingual [interlingua] approach forces unneeded processing.”

3). Transfer Approach: Experience with linguistically ambiguous MT systems which used interlingua approach led to the adoption of more modest ‘transfer’ approach. This approach includes three stages:

- a) Transfer SL text to an SL-oriented abstract representation
- b) Transfer the result representation to a TL-oriented abstract representation
- c) Synthesis of TL text (Arnold, 2003).

These three approaches, i.e. direct approach, interlingua approach, and transfer approach, are called rule-based approaches.

Vauquois (1986) represented the rule-based approaches in a pyramid as follows:

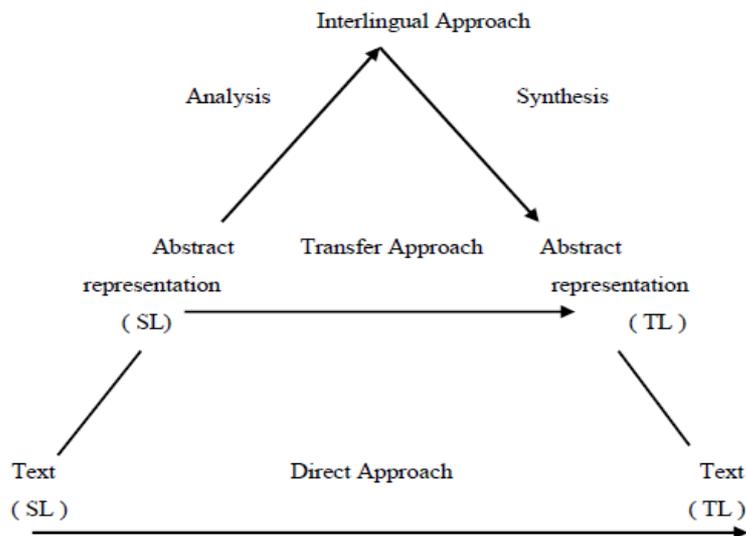


Figure 1. The pyramid of rule-based approaches

1.2 Theoretical Framework: Lexical-Functional Grammar/LFG

“The first was Joan Bresnan, a syntactician and former student of Chomsky’s, who had become concerned about psycholinguist evidence that seemed to show that something was wrong with the concept of transformation [in Chomsky’s theory]” (Dalrymple, 2001: 3). So, she presented an alternative approach in which part of the work done by transformation in Chomsky’s theory was done in the lexicon instead (ibid). Bresnan herself called this approach as ‘lexical-interpretive model of transformational grammar’ in which the terms ‘lexical’ and ‘interpretive’ imply the importance of lexicon and semantics in this model (Dabir Moghaddam, 2004).

1.2.1 Lexical-Functional Grammar’s Levels

In LFG, each sentence has three levels: Constituent Structure / C-structure, Functional Structure / F-structure, and Argument Structure / A-structure.

Dabir Moghaddam (2004) presents a figure of LFG as follows:

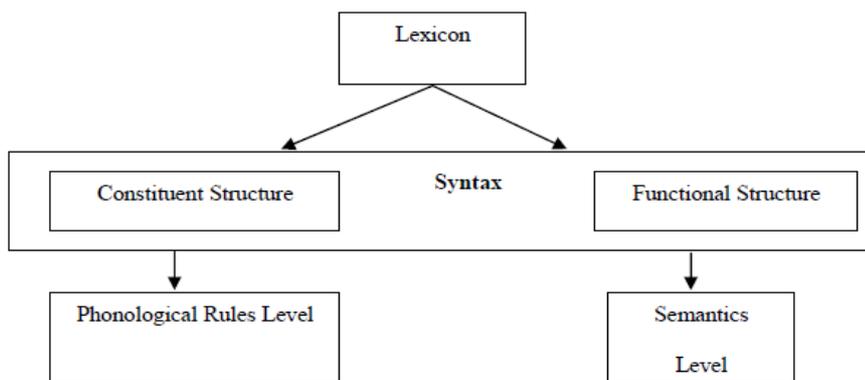


Figure 2. Model of lexical-functional grammar (Dabir Moghaddam, 2004: 519)

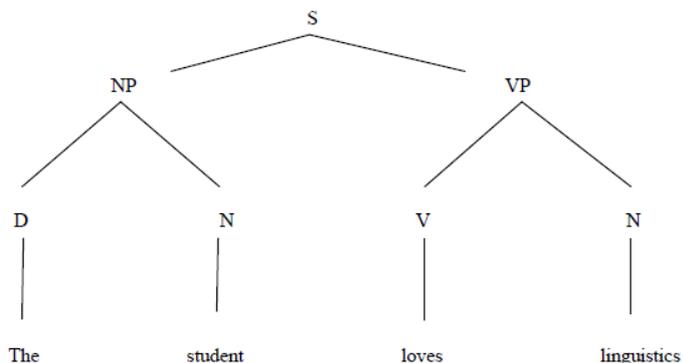
1.2.1.1 C-structure

Carnie (2007) believes that one major part of LFG is almost identical to the transformational grammar. “This is the idea that words of a sentence are organized into constituents, which are represented by a tree, and generated by rules” (Carnie, 2007: 438). Each tree represents C-structure of the given sentence. What makes the tree-diagrams in LFG different from the tree-diagrams in transformational grammar is the fact that in LFG there is no movement and no trace any more (Carnie, 2007).

In C-structure, relations such as dominance, precedence, and constituency are expressed through a series of phrase structure rules, which are represented in tree-diagram (Chatsiou, 2010).

The tree-diagram of the sentence (1), representing its C-structure, is given as the following:

(1) The student loves linguistics.

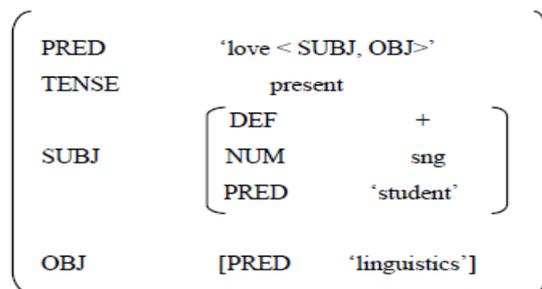


1.2.1.2 F- structure

Another structure in LFG is F-structure. F-structure “represents the relational structure of the sentence” (Van Valin, 2001).

Dabir Moghaddam (2004) believes that in F-structure, not only semantic information of each entry but also grammatical functions of the sentence constituents (i.e. subject, object, and verb) are represented. Moreover, grammatical functions such as subject and object are considered as nominal predicate. Also, for each noun, one ‘specifier / SPEC’ and one ‘number / NUM’ are recorded. Verbal predicate is consisted of a verb and its arguments (such as subject, object, subject complement). In F-structure, grammatical functions are called ‘attribute’ and their correspondence are called ‘value’. So, such a representation is called ‘Attribute Value Matrix / AVM’.

The F-structure of the sentence (1) is as follows:



1.2.1.3 A-structure

Huang states that “C-structure and F-structure are language-dependent, that is they analyze the sentence based on linguistic information (segmental information) present in the sentence. But information related to the upper levels (i.e. supra segmental information), which are not present in the actual presentation of the sentence, should be considered to determine the meaning of the sentence. Thus, A-structure which presents the semantics information of the sentence is also paid attention to in LFG.” (1993, cited in Faroughi, 2007).

Bresnan (1995, cited in Faroughi, 2007: 60) takes two parts for A-structure: Head or predicate and argument. In LFG, A-structure of a sentence shows the number of the participants in an event. Some of these arguments are obligatory and some are optional. This means that obligatory arguments cannot be deleted but optional arguments can be deleted.

The A-structure of *place* is as below:

Place ‘place < (↑ SUBJ) (↑ OBJ) (↑ OBJ_{LOC}) >’ (Dabir Moghadam, 2004: 90)

2. Method

In this part, first of all a description of zero-place predicators will be presented and then the mechanism used in machine translation will be explained.

2.1 Zero-place Predicator

Yarmohammadi (2002) provides a complete contrastive description of zero-place predicator constructions of English and Persian. The patterns of zero-place predicators are as follows:

1). Predicator denotes some point or period in time.

Pro-Subj BE NP
It be NP₁
E.g. (1) It was night.

/ʃæb bud/

(2) It is five O'clock.

/saʔæt-e pændʒ ʔæst/

2). Predicator refers to stretches of time such as day, month, or year.

Pro-Subj BE NP
It be NP₁
E.g. (3) It is April.

/mah-e ʔavrɪl ʔæst/

(4) It is Friday.

/ruz-e dʒomʔe ʔæst/

(5) It was 1995.

/sɔl-e 1995 bud/

3). Predicator refers to some sort of sensation or describes general weather conditions.

Pro-Subj BE Adj
It be Adj
E.g. (6) It is cold.

/hævɔ særd ʔæst/

(7) It is fair.

/hævɔ xub ʔæst/

4). Predicator denotes a weather phenomenon and is concerned with events involving precipitation (a verb (V) in English and Persian).

Pro-Subj Verb
It V
E.g. (8) It is raining.

/barɔn mibræd/

(9) It was snowing.

/bærf mibræd/

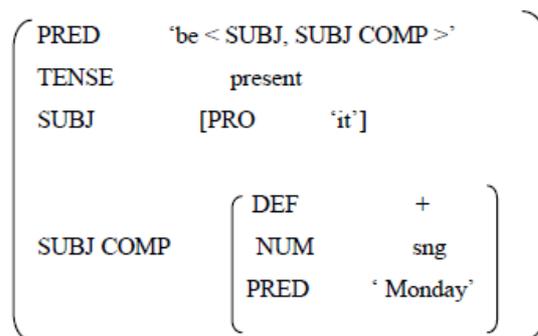
Yarmohammadi (2002) continues that in all the above patterns the element *it* is considered *semantically empty*. But syntactically it functions as filler. The verb *to be* is considered a dummy verb because it does not add any semantic meaning to the sentence. The third elements of these patterns are supposed to be predicators. Therefore, since there is no other element to occupy a position in these patterns, they are called zero-place predicators.

On the contrary, Persian has no dummy *it* to function as subject. English zero-place predicators should be translated into Persian by adding a suitable expression to the beginning of the sentence. So, the corresponding patterns for English zero-place predicators are one-place predicators in Persian (see examples 1-9).

Using a tree-diagram to determine part of speech of words of a sentence is very helpful in cases of lexical ambiguity in machine translation. For instance, 'bark' may be a noun or a verb. If it is used as a verb, it means 'to make a short loud sound by dog' and where it is used as a noun, it means 'the outer covering of a tree'. In such cases, a suitable lexical equivalence can be selected from the dictionary of the machine translation, if the syntactic parser has determined the part of speech of the given word correctly.

Moreover, the determination of grammatical function of noun phrases is essential for producing a natural translation of them into Persian. Thus, an algorithm is written which helped the machine translation for determination of grammatical function of noun phrases present in the sentence based on F-structure.

For instance, the F-structure of the sentence (9) is as follows:



According to the above F-structure, the grammatical function of 'it' is SUBJ and of 'Monday' is SUBJ COMP.

2.2 Using a Dictionary for Selecting a Suitable Equivalence

When the part of speech of words and the grammatical function of noun phrases are determined, it is time to select a suitable equivalence for each word of the sentence. To accomplish this purpose, the word /ruz/ "day", /sal/ "year", /sa?æt/ "hour/clock", and /hæva/ "weather" was added to words which present these concepts (e.g. Saturday, 3 O'clock, and sunny) in the dictionary of the system. In fact, by adding the suitable word beside subject complement, the role of the subject played by the dummy *it* would be played by a meaningful word in Persian sentence.

2.3 Generating Target Language

When the machine translation system determined the grammatical function of noun phrases present in the sentence and selected the suitable equivalence for them from its dictionary, then an algorithm is written for zero-place predicators based on the following sentences:

First, *It* should be deleted;

Second, the suitable equivalence of subject complement should be selected;

Third, the 'to be' verb should be translated.

A part of the given algorithm is as follows:

```
private static void Return2Array(Parse parse, string[] a, string[] b)
{
    for (int i = 0; i < parse.ChildCount; i++)
    {
        Parse hparse = parse.GetChildren()[i];
        Return2ArrayRec(hparse, a, b);
    }
}
private static void Return2ArrayRec(Parse parse, string[] a, string[] b)
{
    for (int i = 0; i < parse.ChildCount; i++)
    {
        if (parse.GetChildren()[i].ChildCount == 0)
            continue;
        if (parse.GetChildren()[i].GetChildren()[0].ChildCount == 0)
        {
            int j;
            for (j = 0; j < a.Length; j++)
```

```

        {
            if (a[j] == null)
                break;
        }
        a[j] = parse.Children()[i].Type.ToString();
        b[j] = parse.Children()[i].Children()[0].ToString();
    }
else
    {
        Parse hparse = parse.Children()[i];
        Return2ArrayRec(hparse, a, b);
    }
}
}

```

3. Results and Conclusion

English zero-place predicators are translated into Persian as one-place predicators and with consideration of Persian word order which are more natural for people whose mother tongue is Persian.

The following sentences were given to the system for translation:

(10) It was night.

شب بود.

(11) It was 6 o'clock.

ساعت 6 بود.

(12) It was Saturday.

روز شنبه بود.

(13) It is June.

ماه ژوئن بود.

(14) It was 2005.

سال 2005 بود.

(15) It is cold.

هوا سرد است.

(16) It was fair.

هوا خوب است.

(17) It is raining.

باران می بارد.

(18) It was snowing.

برف می بارید.

This paper attempted to show that a machine translation system designed based on a lexical- functional grammar model can present a natural translation of English zero-place predicators into Persian. These predicators are structures which do not have a one-to-one correspondence in Persian so presentation of translation for them in machine translation is to some extent difficult.

The findings of the research show that such a machine translation system is able to determine the part of speech of the words and the syntactic category of the phrases of the sentences correctly. Also, utilizing a bilingual dictionary which adds suitable expressions, if necessary, makes the machine translation able to produce a natural translation of English zero-place predicators into Persian, so that it is more comprehensible for Persian speakers.

Furthermore, generating the given sentences based on Persian word order increase the degree of comprehensibility of produced sentences.

The question raised here is that whether this system is able to produce a natural translation of other types of English predicators into Persian.

Acknowledgements

We must express our deep appreciation to Amir Hossein Mohajezadeh who gave us his support by computer programming and algorithm writing throughout this project.

References

- Aasi, M. (2004). Machine Translation: From Dream to Reality. In F. Farahzad (Ed.), *Proceeding of two translation studies conferences* (pp. 31-48). Tehran, Iran: Yalda Ghalam.
- Arnold, D. (2003). Why translation is difficult for computers. In H. Somers (Ed.), *Computer and Translation: a translator's guide*. USA: John Benjamin's North America.
- Bolshakov, I., & Gelbukh, A. (2004). *Computational Linguistics: Models, Resources, Applications*. Mexico: Ciencia de la Computacion.
- Bresnan, J. (1995, October). *Lexicality and argument structure*. The Paris syntax and semantics conference, Paris. Retrieved from <http://www.lfg.stanford.edu/lfg/archive/archive>
- Carnie, A. (2007). *Syntax: A Generative Introduction* (2nd ed.). USA: Blackwell Publishing.
- Chatsiou, A. (2010). *An LFG Approach to Modern Greek Relative Clauses*. PhD dissertation. Department of Language and Linguistics, University of Essex.
- Dabir Moghadam, M. (2004). *Theoretical Linguistics: Emergence and Development of Generative Grammar* (2nd ed.). Tehran: Samt.
- Dalrymple, M. (2001). *Lexical-Functional Grammar*. New York: Academic Press.
- Faroughi, J. (2007). *A Context-based Model for Machine Translation of Some Simple Sentences from English to Persian based on Lexical-Functional Grammar*. Mashhad: Ferdowsi University of Mashhad dissertation.
- Fromkin, V., Rodman, R., & Hyams, N. (2007). *An Introduction to Language* (7th ed.). USA: Heinle.
- Hesabi, A. (2006). Abstract Representation in Machine Translation. *Translation Studies*, 13, 41-48.
- Huang, C. R. (1993). *Mandarin Chinese and the lexical mapping theory: a study of the interaction of Morphology and argument changing*. The Bulletin of the institute of History and Philology.
- Hutchins, J. (1979). Linguistic models in machine translation. *UEA Papers in Linguistics*. Retrieved from <http://www.hutchinsweb.me.uk>
- Hutchins, J. (2003). Machine Translation: General Overview, by Mitkov, Ruslan. *The Oxford Handbook of Computational Linguistics*, 501-511. Oxford: Oxford University Press.
- Khodaparasti, F. (2004). Persian Machine Translation. *Pazand*, 1, 39 - 52.
- Newton, J. (1992). *Computers in Translation: A practical Appraisal*. USA and Canada: Routledge.
- Niknejad, S. A. (2008). *Designing and Building a Machine Translation from Persian into English*. Tehran: Sharif University of Technology M.A. Thesis.
- Saedi, Ch. (2008). *Machine Translation of Persian Sentences into English: A Hybrid Approach*. Tehran: Science and Research Branch Islamic Azad University M.A. Thesis.
- Van Valin, R. (2001). *An Introduction to Syntax*. USA: Cambridge University Press.
- Vauquois, B. (1986). A survey of formal grammar and algorithms for recognition and transformation in Machine Translation. In S. Nirenburg, H. Somers, & Y. Wilks (Eds.), *Readings in machine translation*. Cambridge: MIT Press.
- Wilks, Y. (2009). *Machine Translation: Its Scope and Limits*. UK: Springer.
- Yarmohammadi, L. (2002). *A Contrastive Analysis of Persian and English (Grammar, Vocabulary and Phonology)*. Tehran: Payam Noor University.
- Younesifar, F. (1994). *Designing and Conducting a Machine Translation system based on Syntactic ways*. Tehran: Sharif University of Technology M.A. Thesis.