CHAPTER ONE
INTRODUCTION

1.1. Overview

The present research is a descriptive corpus-based translation study aiming at pinpointing the patterns of translation into Persian when dealing with English Verb Phrase Ellipsis (VPE) and finding out how that observed translation behavior may be taken as advantageous information for improving English-Persian Machine Translation (MT) systems performances.

1.2. Corpus-based Translation Studies

The word *corpus* is a Latin word that signifies "body". It refers to any written or spoken text. However, in modern Linguistics corpus is large collections of electronic and machine readable form texts which represent a sample of a particular variety or use of language(s) (Garner, 2004, p. 226). In this regard, McEnery and Wilson (2001, p. 29) indicate "any collection of more than one text can be called a corpus"; however, in the context of modern linguistics modern corpora have the following characteristics (p. 32): Sampling and Representativeness, finite size, machine-readable form, and Standard reference.

Corpus plays an important role both in Linguistics and in Natural Language Processing (NLP). It aims at providing real, authentic language example, which may constitute a more reliable resource for analyzing linguistic phenomena. The essential characteristics of corpus-based analysis, according to Biber et.al (1998, p. 4), are:

1. It is empirical, i.e. it analyzes the actual patterns of use in natural texts;
2. It utilizes a large and principled collection of natural texts, known as a "corpus";
3. It makes extensive use of computers for analysis;
4. It depends on both quantitative and qualitative analytical techniques.

In recent years there have been many applications of corpora to various disciplines, including translation studies (Xiao and McEnery, 2002; Zhang, 2002; Granger, 2003; Bernardini & Zanettin, 2004; Yan et al., 2008; and Guangrong 2009). The use of corpora in translation and contrastive studies has increased since the 1980´s and the first corpus for translation studies purpose was compiled in 1995 (Baker, 1996). During this time, the methods and tools of corpus linguistics were incorporated into Descriptive Translation Study (DTS) (Toury, 1994). Toury was the first one who elaborated the concept of DTS. As Farahzad (2000) puts it, "studies of this type are able to examine areas such as decision-making in translation, translation norms, and universals of translation" (p. 20). In the early 90´s, Baker (1993, p. 243) predicted that "the availability of large corpora of both original and translated texts, together with the development of a corpus-driven methodology, would enable translation scholars to uncover the nature of translated text as a mediated communicative event." Varantola (2003) states "the knowledge of how to compile and use corpora is an essential part of modern translational competence, and should therefore be dealt with in the training of prospective professional translators." (p. 56)

This study adopts Toury's perspective that any research in translation should start with observational facts (translated utterances and their constituent elements) towards that reconstruction of non-observational facts (Toury, 1994, p.18). By this, the author meant the translated texts are to be studied focusing on the techniques and the strategies the translators have adopted in dealing with, principally, the contrasts of the languages in question. This kind of study helps stating some predictive or explanatory rules about those contrasts and helps achieving a good understanding of the probable errors due to the differences between patterns in the first language (L1) and those found in the second language (L2). Understanding the nature of those errors is necessary before systematic means of resolving them could be found.

In order to define translational behavior in some form of patterns and rules it is necessary to analyze a large quantity of texts which can be considered as a representative of a given phenomenon in a language. Since it is impossible to analyze all the utterances of a language, a sample that can characterize linguistic behavior in that language can be analyzed instead. Due to the fact that analyzing such a large selection of texts manually is not feasible and since this is a tiresome task, highly subject to human error, machine readable texts must be processed using robust (Natural Language

Processing (NLP) tools. The machine-readability facilitates corpus-based studies because the corpora can be searched and manipulated easily and quickly by application of adequate tools which can read the texts, preprocess and normalize them, locate and align all the occurrences of any linguistic pattern along with the corresponding sentences in the other language, extract the collected data, add some other information to the text, for example annotating it with part-of-speech (POS) tags in order to enable researchers to carry out some linguistic investigations in attaining other purposes, and so on.

Different kinds of corpus can be construed depending on their specific purposes. A parallel corpus, i.e. a series of source texts aligned with their corresponding translations, is a good basis for studying translation procedures.

The goal of the present study is to study Persian texts translated from English in order to scrutinize the systematic strategies used in the translation of VPE structures, thus help to identify translation norms. By exploring the probable regularities to be found in these strategies, it may be possible to help define rules for improving the performance of English-Persian Machine Translation (MT) systems. Therefore, an English-Persian parallel corpus will here be used as a starting point to retrieve English sentences where VPE instances are observed and to compare them with their Persian counterparts with the purpose of doing a cross-linguistic contrast.

1.3. Verb Phrase Ellipsis (VPE)

The majority of English sentences include a verb phrase (VP) where an inflected verb appears. It means that, in general, for a sentence to be complete it should contain a verbal constituent. However, sometimes sentences can be found that do not present an inflected verb form yet they are intuitively complete (examples (1) and (2)). In fact, repeated material can be zeroed to avoid redundancy (the part(s) in square brackets indicate the elided element(s)):

(1) *John read the magazine and Mary* [*read*] *the newspaper.*
(2) *John went to school and then* [*John/he went*] *to the mall.*

In example (1), the repeated verb `read´ has been zeroed, whereas, in example (2) both the subject `John/he´ and the verb `went´ have been zeroed. This phenomenon can be

called `verb phrase ellipsis´. It occurs, for example, in providing short answers to queries. For discoursive economy, in question-answer sentence pairs, verbs in the answer sentence are often omitted, when they are the same verb appearing in the question sentence, as in example (3):

(3) *Have you gone to the mall?*
- *Yes, I have* [*gone to the mall.*]

In the answer, the repeated material has been omitted since the discursive situation, in this case the previous sentence (the question), enables the reconstruction of that zeroed material.

VPE in verbal groups forming compound tenses frequently occurs in English, so that what remains in the utterance is just the auxiliary verb, while the main verb is deleted; consider examples (4)and (5):

(4) *I was studying, but he wasn´t* [*studying*].
(5) *Have you been cleaning the house whole day?*
- *Yes, I have been* [*cleaning the house*]

Another case, which can be included in the set of VPE phenomenon, is signaled by certain adverbs or clue-words that indicate repetition (e.g. `also, like, too, again…´); the presence of these elements presupposes the presence of the previous utterance of the material they refer to (examples (6) and (7)):

(6) *I read the journal and Mary* [*read the journal*] *too.*
(7) *She had talked with him and so had Mary.*

In example (6) a complete deletion of VPE has occurred, while in example (7) there is a partial deletion of the verbal group, since only the tense operator `had´ has remained.

In many situations the repeated material can be reduced to a “pro-verb” Gerhard (2005, p. 187), that is, a dummy, semantically void, verb replaces that verb/verb phrase, as in example (8):

(8) *I read the journal and Mary did too.*

Here the auxiliary `did´ occupies the position of the verb; these auxiliaries are thus treated as pro-verbs and require the previous presence of a verb, much like a pronoun requires an antecedent. In other words, they act just like pronouns, with the exception that they take the position of a VP rather than a noun phrase position. In spite of not being an ellipsis as in the previous examples, this case will also be studied here.

Finally, some verb subcategorizes a sentential complement/argument that can be deleted if it has already been stated in the previous discourse (example (9)):

(9) *I apologized the teacher, but he did not want to* [*apologize the teacher*].

In this case, the sub-clause with the verb `apologize´ has been omitted. Because this is a dependent clause of `want´, it is considered not a clear-cut case of VPE, but rather just a matter of repeated argument deletion. Nevertheless, because of the translation difficulties it raises for English-Persian MT, this type of deletion will be included in this study.

1.4. VPE in Linguistic Theory and NLP Framework

VPE is a research subject that gives rise to great interest, whether in linguistic theory or in computational linguistics. This interest arises from its important role in the coherence and cohesiveness of texts. Ellipsis is one of the main cohesive devices (the other devices, according to Halliday and Hasan (1976), conjunction, reference, substitution, and lexical cohesion or synonymy), without which one cannot understand and interpret the relationships between sentences in a text.

One of the most interesting aspects of VPE, much explored in the literature is the ambiguity it produces in discourse (Asher, Hardt, & Busquets, 2001, Schwarz, 2004, Martin & McElree, 2008). This is the case of famous ambiguity between *strict reading* and *sloppy reading* (Ross, 1967). Sloppy reading refers to cases where an ellipsis gives rise to a meaning which differs from the meaning of its antecedent whereas in strict reading the missing elements are interpreted exactly as they are in the antecedents. To elaborate this, consider example (10):

(10) *John loves his wife and Bill does too.*

Under the strict reading, the interpretation *`Bill loves John´s wife´* is considered, while under the sloppy reading the structure is interpreted as *`Bill loves his own wife´*.

Zeroed material should always be assumed to be reconstructable from the surrounding (usually the previous) discourse; nevertheless, this is not always the case. For example in some cases, which Harris (1991, p. 109) has called "information degeneracy", ambiguity is a by-side result of certain reduction procedure. For instance, *`He drove the senator from Ohio to Washington´* is produced by one reduction from *`He drove to Washington the senator who was from Ohio´*, and also by another from *`His driving the senator was from Ohio to Washington´* (p. 339).

While ambiguity is itself interesting as a research topic, this paper will only address unambiguous utterances where VPE occurs. One of the reasons for this option is that people try to avoid ambiguity even if they are not always aware of it or if they do not always succeed in doing so. Secondly, if some utterances are ambiguous to humans, they are also undecidable and it is controversial, for NLP systems, if such ambiguities should be removed from texts, or else, kept as they are, adding to the semantic "texture" or "density" of the discourse (double senses).

VPE, usually treated in the larger context of anaphora, is of high theoretical interest in many language sciences research, particularly in theoretical linguistics, namely by exploring how it is interpreted by humans and how the difficult ambiguous cases should be deal with. For example, Harris (1991) has included reduction as a general constraint of natural language responsible for compacting the discourse by reducing (that is, zeroing or pronouning) redundant material. As said above, the fact that ellipsis asks for a reconstruction procedure from the receptor, makes it the major device to produce the cohesiveness in texts. In order to understand a text, one has to interpret the relationship between different utterances in a discourse. In the case of ellipsis, the discourse becomes interpretable by finding the antecedent of the zeroed element.

Although there are extensive theoretical studies on VPE, to the best of our knowledge, there is apparently no considerable work on VPE in the NLP framework, particularly in translation field. In the context of anaphora computational linguistic studies, for example, other kinds of anaphora have received much more attention and considerable work exists on their resolutions (Lappin, 2005, Mitkov, 2002; Mitkov, Evans and,

Orasan 2002; Mitkov, Boguraev, and Lappin, 2001; Lappin & Benmamoun, 1999; and Lappin, 1996). However, ellipsis as a "zero anaphora" (Mitkov, 2002, p.13) has not benefitted such extensive interest. The reason behind this state of affairs may be that it is usually difficult to spot and characterize the elements that have been deleted from the text and to treat them automatically. Besides, since ellipsis involves the spotting of zeroed material, it is deeply anchored in the very structure of sentences, engaging complex psycholinguistic mechanism, among which syntactic structures are paramount.

Among the different kinds of anaphora types, zero anaphora or ellipsis "may be the most sophisticated variety of anaphora" (Mitkov, 2002, p.12) and constitutes a particularly important challenge to NLP. VPE or "verb phrase zero anaphora" as Mitkov (2002, p.12) calls it, is not an exception. For example, Automatic translation of VPE from English to Persian which seems to be one of the problems Machine Translation (MT) systems usually come across and fail to resolve, has not been studied yet. Since elliptical forms are unique to each language, careful description of the corresponding translational patterns is necessary to find a way to improve MT systems´ performance. To the best of our knowledge, the absence of research on English-Persian NLP studies on VPE is a gap that the present study intends to modestly contribute to explore.

This dissertation is structured in 5 chapters: Chapter one is an introduction to the study; the second chapter consists of a theoretical overview on DTS, corpus-based translation studies, VPE in English and Persian, and VPE in NLP framework. Chapter three presents the methodology of data collection, the applied resources and tools applied here, and the analysis procedure. The fourth chapter provides the quantitative and qualitative analysis of data and a discussion of the findings. Finally, Chapter five summarizes the results of the study and provides some suggestions for further investigation on English-Persian VPE.

CHAPTER TWO
REVIEW OF LITERATURE

This chapter is structured as follows: First, in section 2.2, Descriptive Translation Study (DRS) is defined. In section 2.3, corpus and its various types are shed light. Section 2.4 highlights the role of corpus in translation studies. Section 2.5 elaborates how a parallel corpus facilitates studying a phenomenon of a language. In section 2.6, the phenomena of verbal ellipsis and particularly Verb Phrase Ellipsis (VPE) is introduced; section 2.7 presents ellipsis in Persian; in section 2.8 the related studies on VPE in NLP framework are presented; section 2.9 discusses the role of ellipsis in translation; section 2.10 is about the present English-Persian corpus-based translation studies; finally, section 2.11 provides the conclusion of this chapter.

2.1. Overview

Translation studies can be divided into three domains: 1) Theoretical Translation Studies: explaining the phenomena; 2) Descriptive Translation Studies (DTS): dealing with comparative studies and translatability of phenomena; and 3) Applied Translation Studies: tackling with translator training, translation aids, and translation criticism (Holmes, 2000, pp. 175-182).

The present study takes a descriptive translation study approach. As Toury's (1994) states, any research in translation should start with observational facts (translated utterances and their constituent elements) and then the researcher should proceed towards the reconstruction of non-observational facts. In other words, as Toury states, we can transform translation studies of initial individual texts into an independent discipline with its own methodology, theoretical framework, and procedures of analysis to achieve generalisations about the translation behaviour. Granger (2003, p. 50) illustrates "the generalization derived from empirical evidences can only be valid if based on the study of large collection of texts, not just individual instances." In this

regard, Shlesinger (1998) believes that corpus-based translation studies are an offshoot of descriptive translation studies.

2.2. Descriptive Translation Studies (DTS)

Before the 1970s, translation study was totally introspective, that is, studies were based on one´s own thought process or his/her intuition. They were completely theoretical. From 1970s on, scholars came to recognize the limitations of this type of study. Holmes (2000, p. 173) makes the point more clear when he says:

> Many of the weakness and naiveties of contemporary translation theories are a result of the fact that the theories were, by and large, developed deductively, without resource to actual translated texts-in-function.

Toury (1994) put the focus on finding a methodology for Descriptive Translation Study (DTS). According to him, DTS "is a target-text-oriented discipline consisting of carefully performed studies into well-defined corpuses". As Farahzad (2000) states "studies of this type are able to examine areas such as decision-making in translation, translation norms, and universals of translation" (p. 20). Munday (2001) cites that Toury´s target-text-oriented theoretical model "combines linguistic comparison of Source Text and Target Text and consideration of the cultural framework of the target Text. His aim is to identify the patterns of behavior in translation and thereby to ´reconstruct` the norms at work in translation process" (p. 124).

2.3. Corpus: Definition and Types

Corpus, in modern linguistics, refers to large collections of texts which represent a sample of a particular variety or use of language(s) that are usually stored as an electronic database and are presented in machine readable form (Garner, 2004, p. 226). McEnery and Wilson (2001) describe the characteristics of the modern concept of corpora as follows (pp. 29-32):

A) *Sampling and Representativeness:* In order to study a variety of language it is impossible to analyze all the utterances of that variety; therefore, the best solution would be preparing a sample that can characterize a variety of a language. The selected texts must be authentic and commonly used.

B) *Finite size:* Corpus is a body of text of finite size (unlike *Monitor Corpus*- a non-finite corpus, meaning that to which some other texts can be added and consequently its size may grow).

It is noteworthy that it is the needs of a research project that determine whether or not that size of a sample (number of texts) to be considered as representative of a variety of language. In other words, the size depends on what one wants to do with the corpus.

C) *Machine-readable form:* Machine-readability facilitates corpus-based studies because the corpora can be searched and manipulated easily and quickly; they can also be improved by adding some other information (e.g. part of speech tagging) for attaining different goals.

D) *A Standard reference:* As McEnery and Wilson (2001, p. 32) believe there is often a tacit understanding that a corpus constitutes a standard reference for the language variety that it represents. This presupposes that it will be widely available to other researchers.

2.3.1. Types of Corpora

Corpuses are classified based on different criteria such as the number of the languages involved, the content and the form of a corpus.

Tohidian (2008) classifies corpora based on the number of languages involved: a) *Monolingual corpus* contains texts in a single language; b) *Bilingual corpus* involves two languages; and c) *Multilingual corpus* is a corpus involving more than one language. It is subcategorized into *Comparable corpus*, representing a collection of comparable monolingual corpora (subcorpora) in different languages with the same

sampling frame, and *Parallel corpus*, which consists of texts in one language and their translations into other language(s) and it is usually aligned at sentence or word level. Parallel corpora can be *unidirectional* (the translation from one language into other: Language 1 (L1) into Language 2 (L2), or L2 into L1), *bidirectional* (L1 into L2, and L2 into L1), or *multidirectional* when more than two languages are involved.

However, each kind of corpus is designed for different purposes, in other words it is the objective(s) of a study that help formulate the requirements a corpus must have in order to meet the goals (McEnery and Xiao, 2005a). Comparable corpora, for instance, are used to contrast general linguistic features, and parallel corpora are important for translation studies and they offer specific uses and possibilities.

Having in mind the purposes of the present study, that is, analysing the similarities and dissimilarities between the verbal elliptical patterns in English texts and their Persian counterparts, the use of a bilingual (English and Persian), unidirectional (from English into Persian), parallel corpus meets the aims of this study.

In the following the attempt present briefly the general goals of corpus-based translation studies in order to better frame our study on verbal ellipsis in English and Persian using a parallel corpus.

2.4. Corpus-based Translation Study

The use of corpora in translation and contrastive studies has increased since the 1980´s. As Baker (1998) states, in 1988 several scholars of translation like (2000) indicated their dissatisfaction with the use of introspection in translation studies and bemoaned those approaches which considered translation as idealized, speculative entities. They came to recognize translations as observable facts and the fact that they could be studied objectively. Consequently, some theorists like Baker (1993), Toury (1994), and Holmes (2000) began to search for a whole methodological approach to deal with translations. In this regard, they incorporated the methods and tools of corpus linguistics into DTS. Baker (1993, p. 243) predicted that “the availability of large corpora of both original and

translated texts, together with the development of a corpus-driven methodology, would enable translation scholars to uncover the nature of translated text as a mediated communicative event." It was a new perspective in translation studies. Afterwards, many corpus-based translation studies have been carried out. Aston (1999), for example, studied on the effectiveness of corpora as translation and as learning tools. He refers to the important role of parallel texts, as translation aids as opposed to bilingual dictionaries, in the recovery of correct equivalents. He adds appropriate corpora "make possible more idiomatic, native-like interpretations of source texts and a use of more idiomatic, native-like strategies in target texts." (p. 313). Zhang (2002) demonstrated the role of using corpus for investigating the style of a literary translator. Yan et al. (2008) used corpora in order to discover the Patterns of translation of "fear" Metaphor from English into Chinese. Guangrong (2009) studies manifests that with the help of parallel corpora it is possible to learn some useful strategies to deal with keywords and collocations in translation.

Corpus-based translation studies, according to Hunston (2002), are theoretical and/or practical. By theoretical studies he means the study of the translation process by "exploring how an idea in one language is conveyed in another language and by comparing the linguistic features and their frequencies in translated L2 texts and comparable L1 texts" (p. 123). Xiao & McEnery (2005b), for example, used an English-Chinese parallel corpus to study how temporal and aspectual meanings in English are expressed in Chinese. By a practical approach, it is meant that a corpus is used for developing systems like machine translation and computer-assisted translation systems or for training translators, where the trainees encounter examples of language features in source language (SL) with their aligned translation in target language (TL).

As mentioned before corpus-based translation study is a new perspective in translation studies, particularly when the concerned languages are English and Persian. There are few works using corpora as a tool to study English-Persian translation studies. To the best of our knowledge, they are all M.A. dissertation such as Abedi´s (2004) which compares the thematic organization in two languages of English and Persian;

Shahrestanee (2007) studies units of translation in English-Persian literary Translations; a corpus-based study on translation of the English passive voice into Persian was carried out by Kerdar (2005); and Mohammadian (2007) presented norm-governed strategies of dialogue translation from English to Persian.

The present study, as a novel study, incorporates a parallel corpus to highlight the English sentences with verbal ellipsis and observe how Persian translators rendering these sentences tend to dealing with them, with the purpose of proposing translation patterns (related to verbal ellipsis) based on the patterns the translators have adopted in dealing with English elliptical patterns.

2.5. Parallel Corpus-Based Study

According to Malmkjær (1998), parallel corpus or translation corpus gives a comparative view of characteristics of translated texts, based on which target language norms can be identified. The identification of these norms can be exploited in different fields such as machine translation programs, bilingual dictionaries, language learning/teaching purposes, and translator training. The author adds: "A parallel corpus is a reliable source to investigate differences and similarities of the languages under question" (Malmkjær, 1998, p. 539).

Schäffner (2000) also states one of the main applications of parallel corpora, the determination of text typology:

> Creating an appropriate translation often means adapting the target text to the text-typological conventions of the target culture. Such knowledge can be gained by a comparative analysis of parallel texts (p. 83).

As the present study aims to shed light on the elliptical constructions present in the parallel corpus, the following section gives a brief definition of ellipsis, and particularly verb phrase ellipsis.

2.6. Ellipsis: Verbal Ellipsis

According to Halliday and Hasan (1976), ellipsis is a substitution by zero; it refers to those cases in which there is a structural gap that can be completed by reference to a related structure in the immediate text, usually preceding (grammatical/syntactic ellipsis). Ellipsis is applied to avoid repetition. For example, in answer to the question `Who went to the store?´, one may answer `Max went to the store´. In most contexts, however, speakers/writers would employ an elliptical form, such as: `Max went´, `Max did´, or even just `Max´. This is because they express it more economically by leaving off part(s) of what is present in the question.

Ellipses, comprehensively classified by Halliday and Hasan (1976), are Nominal, Verbal, and Clausal, along with their sub-categorizations. Mitkov (2002, pp. 13-14) states the most common forms of ellipsis are: zero pronominal anaphora, zero noun anaphora, zero verb anaphora, and verb phrase zero anaphora. Consider the following examples (Ø, a symbol used by Mitkov (2002), indicates the ellipted part):

Zero pronominal anaphora: *Mary opened the door and Ø run out.* (Ø → she)
Zero noun anaphora: *William passed two exams and his brother passed three Ø.* (Ø → exams)
Zero verb anaphora: *William passed two exams and his brother Ø three exams.* (Ø → passed)
Verb phrase zero anaphora*: I joined the army but my friend didn't Ø.* (Ø → join)

The present study focuses on verb phrase ellipsis (VPE), what Mitkov (2002) calls it verb phrase zero anaphora while Halliday and Hasan (1976) call it verbal (lexical) ellipsis.

Verbal ellipsis is an ellipsis within the verbal group. A verbal group contains a lexical element (the lexical verb) and operators (the auxiliary verb) which form the rest of the verbal group, expressing grammatical values such as finiteness, modality, polarity, voice

and tense. Accordingly, verbal ellipsis can be divided into two categories: lexical ellipsis (example 1a) and operator ellipsis (example 1b).

(1) *a. Are you swimming?*
- *Yes, I am* [*swimming*].
  b. *What are you doing?*
- [*I am*] *Swimming.*

Lexical ellipsis or verb phrase ellipsis (henceforth VPE) is an instance of verbal ellipsis where the lexical verb is omitted and the presence of an auxiliary verb helps recover the deleted verb.

In an operator ellipsis, the lexical verb is always present while the grammatical elements of verbal group (operators) are missing. In an operator ellipsis, the subject is also omitted.

2.6.1. Verb Phrase Ellipsis (VPE)

As mentioned above (section 2.6) VPE is an instance of ellipsis where the main verb is omitted and what remains is an auxiliary verb.

In this study, Halliday and Hasan's (1976) classification of VPE is presented based on the kind of auxiliaries (i.e., be, have, do, and modal verbs) in the verbal group and the approach is based on Lobeck´s (1995) reference to the constructions in which VPE usually occurs.

An auxiliary verb has one or more of the following syntactic functions: passive, progressive, perfective, modal, or dummy.

A) `Be´ and `have´: `be´ and `have´ with their all paradigms can act either as auxiliary or as light/lexical verb, such as *`He is* (auxiliary) *driving fast´* vs *`He is* (light verb) *a doctor´* and *`He has* (auxiliary) *bought a car´* vs *`He has* (lexical verb) *a car´*.

B) The dummy auxiliary `do´ is used to form interrogatives (e.g., *`Do you like her?´*) and negative structures (e.g., *`I don´t like her´*.) and negatives. This auxiliary can appear alone where a main verb has been omitted, leading into a verb phrase ellipsis or so-called pro-verb; in fact, the verb `do´ acts as a pro-verb to avoid repetition, as in example (2):

(2) *I never wake up soon, but my mother does*. (`does´ is the pro-verb of `wakes up´.)

C) There are nine modal verb forms: `can, could, may, might, will, would, should, must´, and `have to´. They differ from the other auxiliaries because they have both syntactic function and semantic value, and also they cannot function as main verbs.

Lobeck (1995) (cited in Busquets, 1997, pp. 6-7) summarizes the general features of VPE in English. Accoerding to this author, the syntactic environments in which VPE usually occurs are as follows:

- VPE usually, but not always, occur at the end of sentences, as in example (3).

(3) *John came to dinner, but his brother didn´t* [*come*].

- VPE can occur in a coordinate or subordinate clause other than that containing the antecedent (example (4)), and in parallel discourse relations (example (5)).

(4) *Mary hates John and Sue does too.*
(5) *Mary hates John but Sue doesn´t* [*hate John/him*].

- VPE can occur in dialogues, as in example (6):

(6) A: *Does Peter want to go to Toulouse.*
B: *Yes, he does*. (That is, `he wants to go to Toulouse.´)

VPE, according to Lappin and McCord (1990), can also follow a "bare occurrence of the complementizer *to*" (p. 202), as in example (7):

(7) *John sent the email to everyone who asked him to* [*send the email*].

Other syntactic environments for VPE occurrence are comparative noun phrases (NPs) (example (8)) and adverbial clauses (example (9)).

(8) *John ran faster than Mary did.*
(9) *John reached the final point before Mary did*.

2.7. Ellipsis in Persian

Reviewing Persian literature, one can see that only a few instances of ellipsis have been referred to in the grammar books[1]. The Persian grammar writers have neither discussed it broadly, or categorized it perfectly nor talked about the differences, structures, regularities and mechanisms of these elliptical instances. They have treated Persian ellipsis fairly well in the sense of cohesive Persian texture, i.e. how ellipsis as one of the cohesive ties helps in coherence of text. According to Yarmohammadi (1995) a text has some features which make the texture of a text. One of the textual features is cohesion and one of the cohesive ties is ellipsis. The Persian grammar writers have not focused primarily on ellipsis.

Noormohammadi also (1988) in his thesis 'A contrastive analysis of cohesion in English and Persian' focused chiefly on cohesion and dealt with ellipsis not extensively enough. He analyzed four short texts of one page; he consequently encountered few numbers of examples of ellipsis to discuss over them. Toosarvandani (2009, p. 60) focuses on Persian V-Stranding verb phrase ellipsis in complex predicates (light verb constructions) where the head verb is omitted and what remains is the light verb (this kind of VPE is also included in the present study). The following example has been extracted from his paper (p. 62):

sohrab piranha-ra otu nazad, vali rostam [*piranha-ra otu*] *zad*
SOHRAB SHIRT/ACC//PL IRON/NC+NEG+HIT/LV/PST-3SG BUT ROSTAM SHIRT/ACC/PL IRON/NC+HIT/LV/PST-3SG

[1] Yarmohammadi 1995, Farshidvar 1997, Anvari and Givi 1999, Shariat 2001, and Khanlari 2001.

Sohrab not hit iron the shirts but Rostam hit iron the shirts. (That is, `Sohrab didn´t iron the shirts but Rostam ironed the shirts.´)

In the above example, the nominal component of the light verb construction اتو (out) `iron´ along with the internal argument پیراهنها را (piranha-ra) [SHIRT/ACC-PL] `the shirts´ has been deleted.

Hashemianjazi (1989), in his study of English and Persian ellipsis, classified ellipsis based on presupposition of the elements, i.e., how translators can recover the gaps caused by ellipsis in source text, according to the norms of the target language. Mohebbi (1998) also presents a classification of Persian ellipsis as a result of a contrastive analysis; he has presented some sentences with ellipsis extracted from English plays and has analyzed how they are supposed to be translated into Persian. However, the translation available has just been produced by one translator, namely by Mohebbi himself. Another shortcoming that can be criticized in his study is that he apparently has only translated the sentences containing ellipsis, i.e. introspection on isolated sentences.

2.8. VPE in NLP Framework

Reviewing the previous studies on ellipsis, one can say that it has always been a subject of interest in both linguistic theory and computational linguistics. This interest results from its importance as one of the major cohesive ties in production of coherence and cohesion of texts. By cohesive ties and conceptual connectivity of text elements, one can understand the relations between the parts of texts, follow the text, and understand it. However, while other discoursive devices have received much attention from the NLP (Natural Language Processing) community, in the general framework of anaphora resolution, much has still to be done in the processing of ellipsis probably because this linguistic phenomenon is difficult to be dealt with automatically since it is not an easy task to spot the elements which have been deleted from the text.

In the context of computational linguistics, the resolution of other kinds of anaphora, rather than zero anaphora (or ellipsis) has been one of the centers of attention. Many

scholars taking different approaches ranging from ´knowledge intensive and inference-based techniques´ to ´statistical modeling´ and ´machine learning´, worked on anaphora for the purpose of text interpretation, dialogue management, query answering, and machine translation. For example, Lappin and Leass (1994) implemented machine learning for identifying the antecedents of third person pronouns and lexical anaphors (reflexives and reciprocals); Kennedy and Boguraev (1996) presented a modified and extended version of that developed by Lappin and Leass, the algorithm they proposed did not require syntactic parser; Mitkov (1995) and also Kennedy and Boguraev (1996) combined linguistic method with statistical approach in anaphora resolution; Ge et al. (1998) worked in a statistical framework; Mikov (2002) took a corpus-based approach relying on linguistic knowledge.

However, zero anaphora has not benefited so much attention as the other types of anaphora. As for VPE, to the best of our knowledge, only some few works can be cited (see De Vries (2009) for a relatively recent overview); they focused on automatically detecting verb phrase ellipses, identifying their antecedents, and resolving ambiguities.

In the following sections, these researches are expanded under two approaches of syntactic and semantic.

2.8.1. Syntactic Approach

Lappin and McCord (1990)**,** Hardt (1997), and Nielsen (2005) took a syntactic approach in resolving VPE. By `syntactic approach´ it is meant resolving some types of ellipsis at syntactic level, by simply copying the antecedent syntactic material to the gap resultant of the ellipsis. In the following, these three studies are presented:

1) Lappin and McCord (1990): They presented an algorithm for resolving VPE in the framework of Slot Grammar (McCord, 1990). The algorithm operates on the output of a Slot Grammar parser (McCord, 1990) and runs in Prolog. It implements an S-structure (parsed surface structure) analysis of VPE in order to resolve the ellipsis. In Slot

Grammar, every word is associated with a symbol (slots) which represents grammatical relations in two ways. The constituents of a phrase and the head word are obtained by filling slots associated with the head. The lexical entry for the word specifies a set of *complement* slots, corresponding to arguments of the word sense, and the grammar specifies a set of *adjunct* slots for each part of speech (Lappin and MaCord, 1990, pp. 197-198). The slot filler rule gives conditions on the filler phrase and its relation to the higher phrase.

S-structure parse representation identifies the relation between the head V and selected arguments and adjuncts of a structured empty elliptical VP on one hand, and the head V and the corresponding adjuncts of an antecedent VP on the other hand; then the algorithm copies the head V of the antecedent VP into the position of the head of the elliptical VP, and specifies which arguments and adjuncts of the antecedent A are inherited by the elliptical V (Lappin and MaCord, 1990, p. 199).

The algorithm identifies VPE by the presence of an auxiliary verb or the infinitival complementizer "to", when the auxiliary verb or complementizer does not have a realized verb complement (p. 201). A candidate antecedent is defined by finding at least one structural relation between the VPE and the antecedent. A new interpreted VP anaphora tree is generated in order to substitute the VPE with the head V of the antecedent. Then the arguments of the antecedent verb are filtered in order to determine which of them are inherited by the VPE. The combination of the new verb heading the elliptical VP and the arguments and adjuncts it inherits from the antecedent verb provides the interpretation of the elliptical VP (p. 201). The authors didn´t report the recall and precision achieved; they just claimed "the success of the algorithm in providing appropriate representation of VP anaphora cases" (Lappin an McCord, (1990, p. 210).

2) Hardt (1997): He studied automatic VPE detection and antecedent selection. First, the VPEs in a small part of two syntactically parsed corpora were manually annotated and then search patterns were for the VP ellipsis resolution system (VPE-RES algorithm) in order to automatically detect the VPEs in the corpora. The automatically found cases of

VPE were compared with the ones manually found. The recall and Precision were 44% and 53% respectively. Afterwards, the antecedents of the detected VPEs were annotated manually. VPE-RES algorithm took all VPs appearing in a context of 3 sentences (the sentence containing VPE and 2 preceding sentences) as the candidates for VPE antecedent. The proper antecedent was selected based on the highest score the candidate antecedent received for possessing the most number of employed syntactic features of recency, clausal relation, parallelism, and quotation". That is the candidates were scored according to the aforementioned features and the highest scoring antecedent was selected as the probable correct antecedent. With comparison with manual antecedent selection[2] the achieved F-score using only recency as the baseline syntactic feature was 76%. With combination of the syntactic features the performance was 46% higher than the result with recency only baseline.

3) Nielsen (2005): Nielsen developed a system by using the most commonly used machine-learning algorithms aiming at automatic detection of VPE, identification of its antecedent, and the resolution of the ellipsis. The algorithms were trained on manually annotated VPEs in 2 corpuses: a) British National Corpus (Leech, 1992) of 100 million words tagged with part-of-speech tags and b) The Penn Treebank (Marcus et al., 1994a and Marcus et al., 1994b) 1 million words syntactically parsed). The F-score achieved for the former was %76 and for the latter was %82.

For the antecedent selection, Nielsen (2005) added the following features to those already used by Hard (1997): word distance, sentential distance, as-appositive, polarity, parallel adjuncts, the subject-verb match (the subject of VPE with the antecedent VP), and the scores obtained by Hardt´s VPE-RES algorithm. These features were all tested in combination with the recency feature, considered as the baseline. The F-score achieved only with recency was about 87%. With combination of the syntactic features, the result made a slight improvement in the performance of the algorithm; the

[2] The comparison between the automatically selected antecedents with the manually selected ones was based on Head overlap, i.e. "The antecedent chosen by the system contains the head verb of the antecedent chosen by the human annotator or vice versa. The head verbs don't have to match and the rest of the antecedents may differ." (De Vries, 2009)

performance was 1.73% higher than the result with recency only baseline.

For VPE resolution, the author defined a set of rules that reconstruct ellipted material in the VPE syntactic slot. Only cases of VPEs that could be resolved by mere copying of antecedent material in the ellipsis site were considered. More complex cases like ambiguous or those with no explicit antecedent were ignored. The author reports an F-score of 80.97% on this (limited) VPE resolution task.

2.8.2. Semantic Approach

De Vries (2009) proposed a semantic framework for ellipsis resolution. He believes that those ellipses which cannot be resolved by means of syntactic constraints are dealt with `semantic approach´; that is the ellipsis is resolved based on information from a semantic representation of discourse.

The corpus (1 million words, taken from Wall Street Journal part of the Penn Treebank) was first parsed by syntactic parser C&C (Clark and Curran, 2004) and then parsed by semantic parser Boxer (Bos, 2005). The output was a Discourse Representation Structure (DRS) (Kamp and Reyle, 1993) which was used to identify VPE and its antecedent. DRS refers to the discourse entities (agents, events, modifiers …) and semantic relations among these entities. This structure contributes to resolve anaphors. As for VPE, one can say that the event has been omitted and for its resolution one can copy the correct antecedent event along with its associated DRS-relations to the ellipsis site.

As for scoring the candidate antecedents he combined Hardt´s VPE-RES scoring mechanism and Nielsen´s genetic algorithm for two reasons: A Genetic Algorithm performs better with continuous values and according to Nielsen (2005) selecting antecedents with VPE-RES gives better results than selecting antecedents produced by a ML algorithm. The candidate antecedents were scored based on the number of semantic features and syntactic features they carried. The features were recency, sentential

complement, semantic similarity of VPE and antecedent subjects, parallelism of propositional phrases, tense, polarity, and modality match.

De Vries (2009) method showed no improvement over baseline performance. He gave some possible reasons for this result such as a) the size of the corpus was small, b) C & C parser produced some errors, and c) the Boxer parser showed some deficiencies in correctly identifying VPE and antecedents.

The two forgoing approaches are not extended to translation field. The present study focuses on presenting the patterns of translation of English VPE into English. Since the main concern of this study is translation and the processes which are expected to help the improvement of Machine Translation (MT) performances, this section is devoted to discussing the role of ellipsis in translation.

2.9. Ellipsis in Translation

The structure of elliptical utterances and the ways of realizing them may vary from one language to the other and problems arise due to the non-correspondence between different languages. The important point here, from the point of view of translation studies, is that both source text (ST) and target text (TT) are produced for a reader whose needs and expectation are match with the norms of his/her language. The translator as a text producer should concern the conventions and norms of the target language.

The judgment that text producers make about what is expected from text receivers often exert influence on the form a text will take. Since two languages may not share the same structural pattern of ellipsis, the translator cannot retain the source text pattern in the target text; therefore, a different pattern matching with target text norms should be adopted.

Human translators usually have no serious problem in dealing with ellipsis in source text (with the exception of very difficult ambiguous cases), because they intuitively understand the meaning of elliptic sentences, so that they can recover easily the deleted

material and fill in the missing words in their translation. MT systems, on the other hand, require some predefined information be available. As De Vries (2009) states this information should be collected and provided for MT systems, otherwise the resultant gaps lead into translation failure. English-Persian MT systems, for example, fail resolving ellipsis, as in example (10):

(10) *Did you go to the cinema yesterday?*
- *No, I didn't* [*go to cinema yesterday*].

GT[3]
آیا رفتن به سینما دیروز؟
a) Aya raftan be sinema dirooz?
INT GO/GERAND TO/PRE CINEMA/LOC YESTERDAY
Did going to cinema yesterday?

- نه، من نکردم.
Na, man nakardam
NO I/NOM NEG+DO/PST-1SG
No, not did-I.
No, I didn´t do.

In the forgoing example, Google translator translated the auxiliary verb `did` which has led in a wrong output. A possible appropriate translation is presented below:

HT
(آیا) دیروز به سینما رفتی؟
a) Aya dirooz be sinema rafti
INTROGATIVE YESTERDAY TO/PRE CINEMA/LOC GO/PST/2SG
a) *Did yesterday to cinema went-you?*
a) *Did you go to the cinema yesterday?*

- نه، نرفتم.
b) Na narafatm.
NO NEG+GO/PST/1SG
b) *No, not went-I.*
b) *No, I didn´t go.*

As it is noticed, the ellipsis has been filled by the antecedent verb نرفتم (naraftam) [NEG+GO/PST/1SG] `not went-I´ (I didn´t go). In Persian, there is no ellipsis of single verbs.

[3] In the example (10), the English sentences have been translated into Persian by Google Translator: http://www.google.com/language_tools?hl=en

2.10. Conclusion

In order to find ways to improve the quality of machine translation (MT) systems, it seems highly significant to discover the factors that influence. One of the factors that proves to be influential in the performance of MT systems is the linguistic phenomenon of ellipsis. In the case of current English-Persian MT systems, such as Google, the tendency towards phrase-by-phrase translation stumbles on this pervasive linguistic phenomenon, since Persian grammar requires VPE resolution and explicitness.

Reviewing the state of art, it is concluded that there is no empirical study (as far as we know) on the impact of the elliptical construction on English-Persian machine translation. Thus, this study aims at investigating elliptical patterns and translation strategies in two languages of English and Persian. Taking in mind the Bakers´s (1993) statement that translation studies needs to incorporate corpus linguistics methodology into DTS so that one can theorize powerful generalization, and taking into consideration parallel corpora is useful for identification of translation patterns and target language norms, the present study takes a corpus-based approach and exploits an English-Persian parallel corpus to attain the goals of the study.

CHAPTER THREE
METHODOLOGY

3.1. Overview

In this chapter, the methodology used in the present corpus-based, descriptive study is presented: in section 3.2, the corpus which is the base of the study, is described, in particular, those aspects most relevant to the current study. In section 3.3, the typology based on which this study is carried out is introduced. Section 3.4 presents the tools used to process and align the corpus, to obtain the patterns, to automatically detect the cases of studies in the corpus, and so on. Section 3.5 provides the data collection procedure. Section 3.6 is devoted to the collected data.

3.2. Corpus

For this study, a bilingual, unidirectional, English-to-Persian was required. Finding parallel corpora for Persian, since this is a resource-poor language, with shortage of digitally stored materials has been proved to be very difficult. Thus, there was not much choice. At the time this study began, only two English-Persian parallel corpora were available: one by Mosavi-Miangah (2009) and the other by Pilevar (2010).

The first bilingual corpus (about 3,5 million words) consisted of individual sentences, rather than a (cohesive, coherent) text. Therefore, it did not show a sufficient number of instances of verb phrase ellipsis, since ellipsis is a cohesive device that is mostly used as a discursive device to join sequences of sentences, and most frequently occurs in closely tied rejoinders.

The second corpus (Pilevar, 2010) consisted of movies subtitles. Its size was deemed relevant to deduce some representativeness of the linguistic phenomena related to VPE,

especially because of the prominently oral and dialogue nature of its content. In fact, ellipsis most commonly occurs in informal conversation (McCarthy and Carter, 2001; and Bonsignori, 2007). In this regard, Bonsignori (2007) states that ellipsis is a typical linguistic feature of spoken language, rather than written language. As some examples of spoken texts he refers to dramas, movie scripts, and transcriptions of spontaneous conversations. Ellipsis occurs less frequently in other types of texts, because the aim is to avoid ambiguity and misinterpretation that may occur. According to this author, legal texts, for example, mostly contain the short sentences, and the lowest proportion of zero pronouns (Bonsignori, 2007). Review of the above, Pilevar´s corpus was considered adequate for the purpose of this study.

The corpus consisted of 1,600 movies´ subtitles[1], thus presenting texts of informal conversations. It contained 612,086 parallel sentences, and about 4 million words in each language.

Though there was no report on the competence of the translators, having inspected some randomly chosen parts of corpus, the resulting subtitles were considered of enough good quality in view of English VPE human translation to Persian. It is due to the fact that human translators usually have no difficulty in dealing with ellipsis as they have base grammar knowledge and satisfactory understanding of the elliptical constructions both in Persian and English.

Although the general quality of the translated subtitles was good, some textual (mostly spelling) problems were found that hindered the detection of VPE. In order to improve the recall of the search pattern used for this task, the English text was pre-processed and some faulty spellings were normalized. The aforementioned textual problems are briefly presented below:

i. No capitalization information: The English text lacked capitalization information,

[1] It has been reported that the corpus has been extracted from www.opensubtitles.org and the movies with subtitles have been provided by Tiedemann (2007) (Pilevar, 2010).

which plays an important role in detecting name entities (person, organization, location...). Consequently, it was not possible to define search patterns including candidate proper names (e.g., `Carl´ in example (1)) that might act as the subject of a sentence. As a result, only subject pronouns (`I, you, she, he, it...´) could be used in the search pattern, to match the subject of verbs involved in VPE.

(1) *we may survive but carl doesnt* [*servive*].

ii. Some lack of punctuation: There was no punctuation (full stops, commas, question mark, colon, or semicolon) in some sentences (in example (2) `this´ and in example (3) `dont´ should be followed by a comma) and in some others the punctuation applied was not appropriate, for instance, a full stop being used instead of a question mark (example (3)) should have ended with a question mark, rather than a full stop, and in example (2) after the tag question `do you´, there should have been a question mark). Lack of question mark did not allow the spotting of the VPEs occurring in tag question forms. Lack of commas, splitting the short answer from the rest of discourse, or lack of full stops (sentence-splitting punctuation) does not allow pinpointing the VPEs preceding them and causes the search patterns to miss some instances of VPE appearing at the end of sentences. For instance, the presence of a comma or a full stop after `don´t´, in example (2), would help to detect the VPE preceding it; that is, the sentence should have been punctuated like `I don´t, they may go´ or `I don´t. They may go´.

(2) *You dont believe any of this do you [believe any of this].*
(3) *Who goes to the concert.*
- *I dont [go to the concert] they may go.*

iii. Bad or wrong spelling: The corpus contained many spelling errors. Hence, it was decided to pre-process and normalize some of the spelling errors in the English text that might have a direct negative effect on the performance of the tool application of the search pattern.

One of the problems resulted because of two different types of contraction:

a) *Auxiliary verb + `not´:* In the combination of auxiliary verbs with negation element (*not*), different forms were observed, e.g., `is not´, `isn´t´, and the incorrect spelling `isnt´. These were preprocessed in to the form `is not´.

b) Subject pronoun+ Auxiliary/Light verb: An incorrect contraction resulting from the combination of subject pronoun and auxiliary/light verbs without apostrophe was also formed, such as `ive´ from `I´ve´. These cases were also preprocessed and changed to `I have´.

These last types of contraction, however, did not hinder VPE detection directly, since ellipsis only occurs in non-contracted forms (except if negation is also involved). However, as the antecedent of a VPE may also be affected by these incorrectly spelled contractions, it was necessary to preprocess them before yielding the text to an MT system. However, since the incorrectly spelled contractions were ambiguous with other words (Table 1), the preprocessing was carried out manually throughout the data analysis.

Table 1. Ambiguous, incorrectly spelled, unsolvable Contractions

Word forms	Equivalence(s)
Hell	he will/hell
Hes	he is/ he has
Ill	I will/ Ill
Shell	she will/ shell
Shes	she is
Its	it is/ it has/its
Well	we will
Were	we are/ were

3.3. Typology of English VPE

Based on Halliday and Hasan´s (1976), a classification of English verbal ellipsis was obtained. This was complemented with remarks taken from Lobeck (1995), as well as Lappin and McCord´s (1990). This typology served as the base for building the search patterns applied to the English corpus in order to retrieve instances of VPE. These

sources are briefly presented below:

According to Halliday and Hassan (1976), VPE occurs in sentences with auxiliary verbs and no main verb. Auxiliary verbs take the position of main verb and lead into VPE; they are passive or progressive (`am, is, are, was, were´), perfective (`have, has, had´), modal (`can, could, may, might, will, would, should, must, and `have to´), or dummy (do, does, did).

Lobeck (1995) refers to linguistic constructions in which English VPE usually occurs:

- Phrase-finally
- Coordinate or subordiante clauses other than that containing the antecedents
- *Yes*/*No* short responses.
- Comparative sentences

To the above list, the Lappin and McCord ´s (1990) reference to the occurrence of VPE after complementizer `to´ was added.

3.4. Tools

3.4.1. Unitex

Unitex[2] (Paumier, 2008) is an open-source corpus processing software; a user-friendly tool that is applied for pattern matching with regular expressions or finite-state graphs. It allows defining search patterns, which are directly built with a visual interface. Because it uses Unicode it can also deal with Arabic-script-based languages, such as Persian.

This tool was used to: (i) normalize the English corpus; (ii) build search patterns in the form of finite-state graphs, apply them to the corpus and retrieve candidate VPE instances; (iii) obtain quantitative data; (iv)build parameterized concordances of matching sequences (Fig. 1); and (v) extract the intended sentences.

[2] http://www-igm.univ-mlv.fr/~unitex/

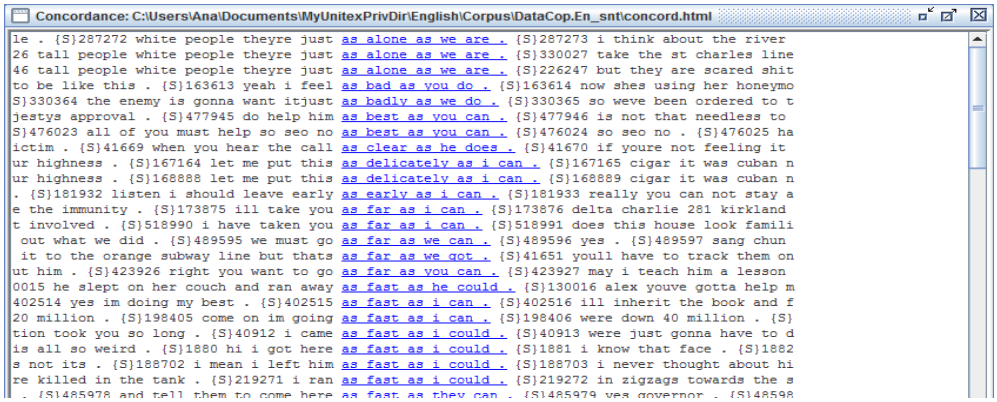

Figure 1: A snapshot of a concordance of the search pattern `as…as´ comparative construction

3.4.2. Python

Python software[3] was used to number the sentences of the English-Persian parallel corpus[4]. This helped extracting the Persian counterparts of the English sub-corpus which in turn contributed in aligning the parallel sub-corpus. Python also aligned Google translation (GT) sub-corpus with the other two sub-corpuses (English and Persian) for the next phase of study.

3.4.3. Translation Tool

From among the free online MT systems, only three contained the English-Persian language pair: Babylon 8, Google, and SDL international. After a preliminary testing, Google translator was deemed to perform better than the other two and so it was selected for this study.

3.5. Data Collection

The parallel English-Persian corpus was used to identify English verbal ellipsis, to establish how VPE is dealt with in translation into Persian, and to try to make generalization on translation behavior in terms of the observed regularities in the

[3] http://www.python.org/
[4] I would like to thank Dr. Vahid Nazari Talooki who helped me in implementing Python for the numbering and aligning of the corpus sentences.

strategies of translation applied by human translators. The goal is to apply this knowledge in the forms of rules in order to improve MT systems output when dealing with VPE.

For this purpose, the first step was to detect the gaps (i.e., the verbal ellipsis) in the English text. The detection was carried out based on the knowledge obtained through studying the Halliday and Hasan's (1976) verbal ellipsis classification and Lobeck ´s (1995) general features of VPE in English. That is, according to this typology of English VPE, a set of search patterns were defined. In the following, this set of search pattern is briefly described.

3.5.1. The Search Pattern

The patterns in which ellipsis frequently occur are, namely, dialogues (in short sequences of responses to `yes/no´ questions, statements, or commands), pro-forms[5] (structures with `so/too´ or `either/neither´), subordinate clauses, and coordinated structures.

VPEs occurring in the middle of sentences are difficult to capture automatically; hence, only those VPEs which appear at the end of the sentences were considered here. The following cases were researched in the corpus: sentences ending with a) auxiliary verbs, b) infinitival complementizer `to´, and c) pro-forms. The search patterns were builts to capture these structures. The general pattern is represented in the graph of Fig. 2. The search patterns were represented by finite-state graphs with the help of Unitex. The words that could be matched were presented in word-forms or POS tags (<ADV>, <ADJ>). These elements are in the nodes of the graph. English auxiliaries were manually given a feature `+aux´ so that they could be represented by the tag <V+aux>. The graphs are read from left to right and they retrieve any pattern that has a path from the initial node () to the end node (). In order to

[5] Quirk et al. (1985, p. 863)) called the structures with `so/too/also´ or `either/neither´, `pro-forms´ (or predication ellipsis), which is the term to be adopted hereafter in this thesis, and Huddlestone and Pullum (2002, p. 787) called them `coordinators´.

simplify the task, only the simplest patterns were considered here, i.e. a clause consisting only of the essential elements, eventually with a facultative adverb as well.

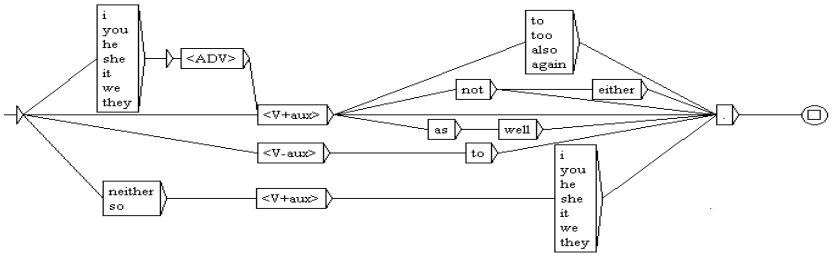

Figure 2. The general search pattern

This pattern constituted the scope of this research. 14,485 cases were captured. The output also included some non-elliptical constructions, unintended VPE (VPEs with compound tenses), and elliptical constructions of different kind rather than VPE. The non-elliptical constructions (2,535 cases, mostly relative clauses (1,078), WH questions (689), the sentences ending to preposition `to´[6] (294 cases), and some errors (474)) and the unintended VPEs (1,435 cases) were discarded auto-manually[7]. There were 2071 instances which were not spotted automatically, whether in form of non-elliptical cases or the elliptical constructions which were not of VPE type; they were ignored throughout of data analysis. The first overview of the remaining data (the intended VPEs, other kinds of ellipsis[8], and some other non-elliptical cases that could not be captured automatically); the resulting matches 10,515 cases) and the distribution of these patterns are presented in table 2; the patterns themselves are described below. All the instances of type (b) and (c), namely 665 and 191 instances, respectively, were studied here (the subjects of all were only pronouns). For the largest type of matches (9,659 cases), 1477 instances were extensively studied (subjects of all were only pronouns); they were restricted to certain linguistic contexts such as: *yes*/*no* short responses to questions; the clauses after coordinating conjunctions `and/but´ and after adverbial conjunctions `after/before´; conditional clauses; and some comparative structures. The rest of the

[6] The search pattern was defined to pinpoint the VPEs appearing after complementizer `to´, the sentences ending with preposition `to´ were also captured (e.g. *I´m not the one you should apologize to).*
[7] After observing repeated cases of unintended structure, based on which a new graph was defined, applied to the corpus, and the cases extracted from the corpus.
[8] The ellipses after verb `be´ which do not belong to VPE; they were ignored throughout the study.

output (8,182) consisted of different subordinate clauses or closely tied rejoinders either in form of a response to a statement confirming or contradicting it or a response to a command obeying or refusing it. From this 8,182 cases, all the instances of VPE with auxiliary `have´ and about 50% of the occurrences of VPE with modal verbs and each of the other two kinds of operators (`do´ and `be´), from three random locations of the corpus (up, middle, and bottom) were studied (in total 4,288 cases: 1,707 cases in short responses to an immediate statement and 2,581 cases in different subordinate clauses) (table 2).

Table 2. The first overview of the resulting matches from the corpus

VPE in clauses ending with	Number of matches		Aux.	Number of occurrences	Studies instances
auxiliary verbs	9,659	Subordinates & coordinated (1,477)	do	436	436
			be	675	675
			have	152	152
			modals	214	214
		Others (8,182)	do	2,034	1,000
			be	2,969	1,500
			have	488	488
			modals	2,691	1,300
complementizer `to´					665
pro-forms					191
Total					6,621

The patterns are described below:

1) A verbal group in which complementizer `to´ occurs at the end and the verb is not an auxiliary (Fig. 3).

Figure 3: Verbal group ending to infinitival complementizer `to´

665 instances were captured (e.g. *I would better see him, I don´t really want to.*)[9].

2) Pro-forms `so/too/as well/also/neither/either´.

[9] Examples are from the corpus.

For pro-forms, two patterns were built: the first (Fig. 4) for `too´, `as well´, and `either´ at the end of the sentence (e.g. *I know how we got so lost and you do too.)*, and the second (Fig. 5) for `so´ and `neither´ (e.g. *we have to interview every applicant for every job and so do you.)*, with inverted subjects.

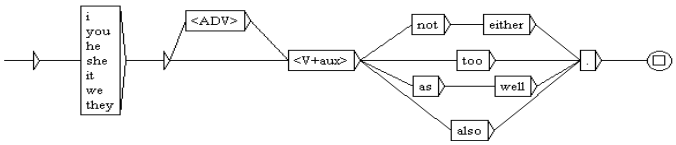

Figure 4: Pro-forms `too´, `as well´, `also´, and `either´

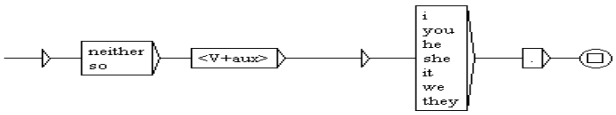

Figure 5: Pro-forms `neither´ and `so´

Due to the problems of the corpus represented in section 3.2, the VPEs with proper names as their subjects were excluded.

The search pattern captured 191 cases, in total. 16 cases were irrelevant; the reason of their presence in the output are rerpresented in section 3.5.1.2.

3) Clauses ending with auxiliary verbs, such as:

a) Clauses appearing after coordinating conjunctions `and´ or `but´ (Fig. 6).

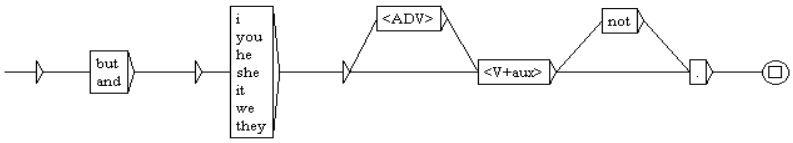

Figure 6: Coordinated Structures (Coordinating Conjunction `and´ & `but´)

308 instances were captured (e.g. *I liked his reaction but he didn´t.)*; among of which 185 cases were irrelevant (e.g. *Thanks for what you did).*

b) Short sequences of responses to `yes/no´ questions (Fig. 7).

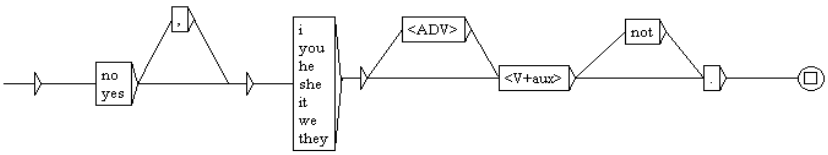

Figure 7: Short responses to `yes/no´ questions

620 instances were spotted, 244 of which were irrelevant cases (e.g. *He seems quite intelligent. Yeah, he may be*).

c) Comparative constructions after the comparative operator `than´, `as´ + Adjective/Adverb + `as´, or `like´ (Figs. 8 & 9).

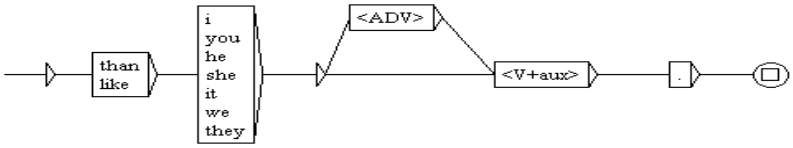

Figure 8: Comparative construction with `than´ or `like´

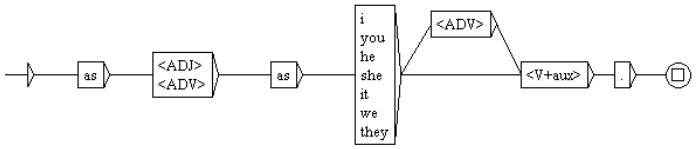

Figure 9: Comparative construction with `as…as´

The numbers of captured cases were: 133 instances with `than´ (43 irrelevant cases; e.g. *This sound is more real than it normally is*.); 38 with `like´ (2 irrelevant case; e.g. *You were damn stupid just like I was*.); and 116 instances with `as…as´ (59 irrelevant cases; e.g. *You are as much responsible as we are*).

d) Adverbial clauses with adverbial conjunctions `after/before´ (Fig. 10).

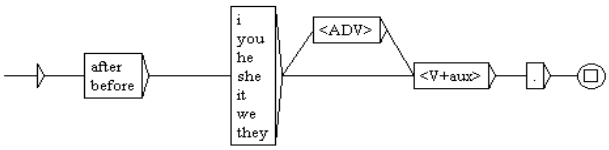

Figure 10: Adverbial Clause (`after/before´)

22 relevant (e.g. *we need to go over a couple of things before you do*) and 1 irrelevant

cases (e.g. *May be he was already on the island before we were*) were spotted.

e) Conditional clauses (Fig. 11).

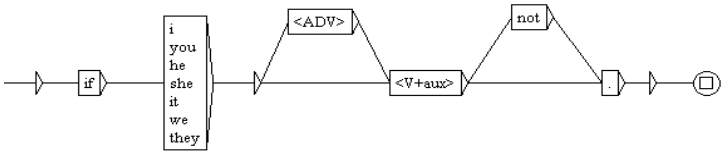

Figure 11: Conditional Clause

156 relevant (e.g. *I just want to catch you if I can*.) and 83 irrelevant cases (example (4)) were pinpointed.

(4) *He may be sleep.*
- *I shall wake him up if he is.*

3.5.1.1. Exclusions

The followings instances of VPE were excluded from the research:

1. *Tag questions*. Since tag question structures end with question mark and the corpus has not been punctuated with punctuations other than full stop, it was not feasible to spot them in the corpus.

2. *Compound tenses formed of more than one auxiliary*. Since the verb that has undergone ellipsis *can* be a simple word (example 5a)[10] or a compound string of auxiliaries (5 (c & d)) the search patterns were limited to strings of up to one auxiliary. Therefore, cases like (5c) and (5d) were excluded. Considering the following examples, when ellipsis occurs, what is retained is the first auxiliary verb, no matter how long the string of auxiliaries is.

(5) *a) John swims well but Mary doesn´t* [*swim well*].
b) John has swum well but Mary hasn´t [*swum well*].

[10] Example (5) is not from the corpus.

c) John has been swimming well but Mary hasn´t [*been swimming well*].
d) John will have been swimming well but Mary won´t [*have been swimming well*].

Besides, it has been noticed that a string of auxiliaries less frequently occurs in a text than the simple ones; thus, for the time being the focus will be on the simple cases. An attempt to find any good solution or method for the improvement of the MT dealing with verbal ellipsis of one auxiliary may also be expanded to incorporate the more complex cases with more than one auxiliary, or it can be the first step contributing the future studies for the more complex cases.

3.5.1.2. Spurious Matches

The defined search pattern also captured some irrelevant cases. For instance, in case of VPE appearing after complementizer `to´, other cases (97instances), ending to preposition `to´, were observed, like *I´m not the one you should appologize to*.

Some other cases included various paradigms of `be´, `do´, and `have´ which do not correspond with the intended pattern. The search patterns were intended to pinpoint various paradigms of `be´, `do´, and `have´ acting as an auxiliary verb (operators); however, some non-elliptical constructions (examples (6), (7), (8), and (9)) or other ellipses rather than VPE (example (10)) were observed in the output. Consider examples (6) and (7); they were captured because `did´ as a pro-verb and `is´ as a light verb have appeared at the end of indirect interrogative sub-clauses; example (8) was captured because `has´ signifying `possess´ has come at the end of a relative clause; example (9) is a direct WH question in which the pro-verb has occurred finally. And finally, example (10), which is of the highest proportion of the irrelevant cases, is an elliptical construction but the elliptical element is clearly an adjective not a verb.

(6) *Thanks for what you did.*
(7) *We don't even know where he is.*
(8) *It´s the only property which he has.*
(9) *What the hell did I do?*
(10) *I am strict and she is* [*strict*] *too.*

3.5.2. English, Persian, and English-Persian-GT Sub-corpuses

In the next phase of the study, the search pattern was given to the tool Unitex to automatically detect the instances of VPE in the English corpus. Unitex provided a concordance of all discovered instances of VPE. It is worth mentioning that, after reviewing the list of the concordances, it was perceived that the search pattern needs to be improved in order to remove some deficiencies in the acquired output. Consequently the process took a form of repetition; that is, after revision, a new and improved search pattern was provided and was given to the tool to investigate the output improvement. It was repeated until the best improved output (the search patterns presented in section 3.5.1) was obtained.

The sentences presenting instances of VPE were stored and they provided the English sub-corpus for this study. In case that one extracted sentence did not contain both VPE and its antecedent, the sentence covering the antecedent, which is usually the one preceding the sentence with VPE, was extracted manually and was added to the output.
In order to study how the English sub-corpus is treated in Persian translation, it was necessary to obtain Persian sub-corpus. Persian sub-corpus was attained through extracting the Persian translation of the English sub-corpus from the original Persian corpus. That is, first the parallel sentences of the preprocessed and normalized English-Persian corpus were numbered by means of Python software, according to the rule shown in Fig. 12: Python 1.

```
#! python
#This code get a file as an input,
puts line number at the first of each line of file
#and produce final output file.

inputstring='/root/corporas/DataRefinedFa.txt'
outputstring='/root/corporas/RefinedFa.txt'

inputfile = file(inputstring)
outputfile = file(outputstring,"w")

linesnumber=0

for i in inputfile.readlines():
      linesnumber=linesnumber+1
      a = i.split()
      tempstr=str(linesnumber)+" "+i
      outputfile.write(tempstr)
```

Figure 12: Python 1: Line numbering

Afterwards, another programme (Python 2) was provided in order that Python can extract the Persian counterparts of English sub-corpus (Fig. 13). The numbered sentences facilitated the task. Python aligned the parallel texts and facilitated the task of comparison and analysis of the collected data. The goal is to draw generalized conclusions from a collection of specific observations in translation of English VPE, which at the end help in defining rules for improving English-Persian MT systems regarding VPE.

```
#this code receives inputfile1 and inputfile2 as an input.
#note:we can run a query on corpora by Unitex.After processing
corpora based on this query,
#we have a new file which we called it Query file.
#inputfile1 is the Query file.
#inputfile2 is the corpora which contains the line number at
first of each line.
#the output file is like inputfile1 file but has line number
#As an example,the query file (inputfile1) has a sentence like
"neither was he ."The line number of this sentence in corpora
(inputfile2) is 1800.
#So the output file (outputfile ) should have a line like "1800
neither was he ."

inputfile1 = file('/root/corporas/infi')
inputfile2 = file('/root/corporas/DataBack.Fa.txt')
outputfile = file('/root/corporas/concord_Fa.txt',"w")
counter=1
temp= inputfile2.readlines()
for i in inputfile1.readlines():
      en = i.split()
      b=int(en[0])-1
      outputfile.write(temp[b])
```

Figure 13: Python 2: Extracting the Persian sub-corpus

In order to improve the performance of English-Persian MT, it was necessary to figure out MT failure in treating with English VPE. For this purpose the English sub-corpus was translated by Google Translator (hereafter: GT) and the GT sub-corpus was produced.

Again Python was programmed (Fig. 14: Python 3) in order to align the sentences of the three sub-corpuses, in a form that each sentence appeared in a separate line (Fig. 15). Now it was possible to analyze the GT failure in translation of English VPE.

```
inputfile1 = file('/root/mitra thesis/corporas/SubAux1.En.txt')
inputfile2 = file('/root/mitra thesis/corporas/SubAux1.Fa.txt')
inputfile3 = file('/root/mitra
thesis/corporas/SubAuxGoogle1.txt')
outputfile = file('/root/mitra
thesis/corporas/JointSubAux.txt',"w")
a=inputfile1.readline()
while (a <> ""):
      b=inputfile2.readline()
      c=inputfile3.readline()
      writestring = a + b+ c+"\n\n"
      outputfile.write(writestring)
      a=inputfile1.readline()
```

Figure 14: Python 3: Extracting the English-Persian-GT sub-corpu

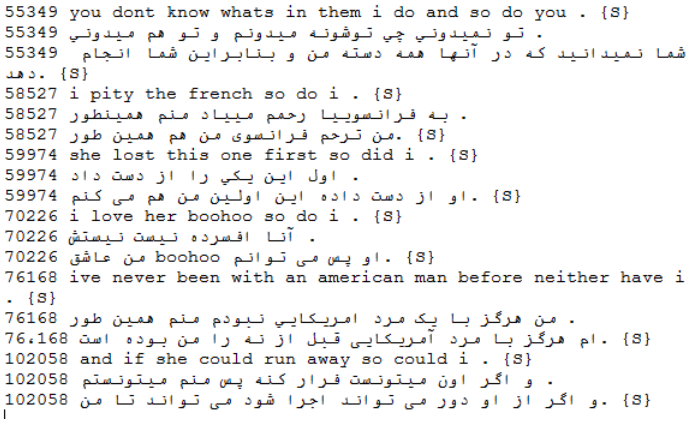

Figure 15: A snapshot of the English-Persian-GT sub-corpus

3.6. Results

6,621 instances of spotted cases out of 10,515 instances were studied, from which 4,523 cases were relevant and 2,098 cases were irrelevant to the study (table 3).

Table 3: Number of VPEs in Different Constructions

Constructions			Studied cases	Relevant cases	Irrelevant cases
Definition	**Pattern**	**Example**			
Subordinate clauses	comparatives	than/ like/as…as	287	183	104
	Conditionals	if…	239	156	83
	adverbial clauses	after & before	23	22	1
	Miscellaneous		2,581	1,648	853
Dialogues	short responses	yes/no	620	376	244
		indirect responses	1,707	1,091	601
non-finite dependent clauses	complementizer `to´	verb + to	665	654	11
coordinated structures	pro-forms	so/too/also/as well either/neither	191	175	16
	Coordinated	but & and	308	123	185
Total			6,621	4,523	2,098

They are ordered according to the linguistic construction to which they belong to. The most frequent cases are VPEs at the end of subordinate clauses (after auxiliary verbs) and in short responses. The least frequent cases are VPEs in pro-forms; VPEs after complementizer `to´ stand in the middle.

Based on the collected data by the analysis of GT failure in dealing with instances of English VPE, and a survey through the strategies adopted by the human translators when

encountered the same instances of VPE, some generalized conclusions were drawn.

In order to verify the representativeness of the obtained results of the study, a sampling is carried out using a randomly- extracted portion of text from 6 different locations of the entire corpus (each containing 5% of the corpus, in total 183,607 sentences). This sub-corpus is then manually analyzed and the spotted instances of VPE are compared against their Persian counterparts in order to assess the translation strategies adopted for them. It should be noted that VPEs after complementizer `to´, and VPEs in pro-forms structure are few numbers in this corpus and they are all studied.

CHAPTER FOUR
DATA ANALYSIS AND DISCUSSION

This chapter is devoted to presenting, analyzing, discussing, and assessing the collected data. The strategies used by Persian translators dealing with elliptical patterns in the English text will be pointed out and discussed. Afterwards, the Google Translation of the English text will be analyzed focusing on Google failure in proper rendering of sentences with VPEs.

4.1. Overview

Table 3, in chapter 3, illustrates the number of VPE instances in different constructions. It was found that the most frequent pattern including VPE (about 79%) consists in subordinate clauses ending in an auxiliary verb, and closely tied rejoinders, either in the form of `yes/no´ answer following a `yes/no´ question, or in form of a response to a statement confirming or contradicting it or to a command, obeying or refusing it. The least frequent pattern consists in the use of pro-form structures and coordinated clauses with conjunction `and/but´ (about 4% and 3%, respectively). About 14% of cases were VPEs after infinitival complementizer `to´.

In this chapter, the structures of VPE mentioned above will be discussed under operators (`do´, `be´, and `have´) and modal verbs, as a guiding line for our text. VPEs occurring in Pro-forms and those after infinitival complementizer `to´ are discussed in separate sections because the structures act differently.

This chapter is, thus, structured as follows: in section 4.2, the VPEs occurring after auxiliary verbs will be analyzed. Section 4.3 is devoted to VPEs after complementizer `to´, and finally, section 4.4 focuses on VPE in comparative pro-forms. The analysis will be discussed through the examples extracted from the corpus. In this chapter, the

examples are presented with no spelling or punctuation errors, which have been observed in the corpus. The next phase of study (section 4.5) assesses the results.

4.2. VPE after Auxiliary Verbs (Operators and Modal Verbs)

In this section, the VPEs which occur at the end of a sentence, after an auxiliary verb are analyzed. As mentioned before, this type of VPE is found in subordinate clauses or in closely tied rejoinders like short answers to questions, statements, or commands. The collected data consisted 5,765 cases (1,436; 2,175; 640; and 1,514 with auxiliaries `do´, `be´, `have´, and modals, respectively) from which 2,071 cases were irrelevant to the study (Table 1).

Table 1. Number of VPE after auxiliary verbs (relevant & irrelevant cases)

Aux.	Studied cases	Relevant cases	Irrelevant cases
Do	1,436	1,025	411
Be	2,175	1,054	1,121
Have	640	480	160
Modals	1,514	1,135	379
Total	5,765	3,694	2,071

The auxiliary verbs in English are of two kinds: 1) Operators which carry grammatical functions (tense, polarity, and voice) such as various inflected forms of `do´ (`do, does, did´) used for negation and interrogation; `be´ (`am, is, are, was, were´) forming progressive or passive forms; or `have´ (`have, has, had´) as a part of present and past perfect tense. 2) Modal verbs which carry both grammatical function and lexical function (possessing a semantic value) such as `can, could, may, might …´.

In Persian, there is no auxiliary verb that can carry the systematic features (polarity and tense) of the finite verbs and these functions are fulfilled differently. For polarity, for example, in English, polarity is normally expressed by the negative element, i.e., `not´ (other negative adverbs such as `never´, `hardly´, etc. may occur in place of `not´).

In Persian, negative morpheme is the letter (ن) prefixed to the lexical verb such as نرفت (naraft) [NEG+GO/PST-3SG] `not-went-s/he´ (`he didn't go´), نرفته بود (narafte bood) [NEG+GO/PP+BE/PST-3SG] `not-gone-was´ (`he hadn't gone´), etc. or modal verbs such as

نتوانست برود (natavanest beravad) [NEG+CAN/MOD/PAST-3SG GO/INF-3SG] `not-could go´ (`he couldn´t go´). If the verb consists of more than one constituent (light verb constructions) (Karimi- Doostan, 1997), the negative element is prefixed to the light verb as in کارنکردم (kar nakardam) [WORK/NC+DO/LV/PAST-1SG] `work_not-did-I´ (`I didn't work´). In simple present tense the negative element is followed by tense operator می (mi) as in نمی روم (nemiravam) [NEG+GO/PRS-1SG] `not-go-I´ (`I don´t/won´t go´).

Tense, in Persian, is designated by taking various forms and affixing the lexical verb as in: میروم (miravam) [GO/PRS-1SG] `go-I´ (`I go´) (simple present or future)[1]; or adding tense element(s) to it as in: خواهم رفت (khaham raft) [WILL/MOD-1SG GO/INF] `will I go´ (`I will go´) (simple future), رفته ام (rafteam) [GO/PP+BE/PRS-1SG] ´gone-I´ (`I have gone´) (present perfect) and so on; except for simple past which the form of verb is infinitive such as رفتم (raftam) [GO/PST-1SG] `went-I´ (`I went´) (simple past). Modal operators accompanying lexical verbs can carry tense element(s) as well as in: می توانم بروم (mitavanaam beravam) [CAN/MOD/PRS-1SG GO/INF-1SG] `can-I go-I´ (`I can go´) (present), توانستم بروم (tavanestam beravam) [CAN/MOD/PST-1SG GO/INF-1SG] `could-I go-I´ (`I could go´) (past), and the like. In light verb constructions, the nominal component remains the same in various tenses, whereas the light verb carries tense by taking various forms as in بازی میکنم (bazi mikonam) [PLAY/NC+DO/LV/PRS-1SG] `do-I_play ´ (`I play´) (simple present or future), بازی کردم (bazi kardam) [PLAY/NC+DO//PST-1SG] `did-I_play´ (`I played´) (simple present), بازی کرده بودم (bazi karde boodam) [PLAY/NC+DO/PP+BE/PAST-1SG] `done-was-I_play´ (`I had played´) (past perfect), and so on. The nominal element بازی (bazi) `paly´ does not carry any tense or even any part of a compound tense; only the light verb کردن (kardan) `do´ fully expresses the tense.

Differences between the linguistic patterns of the two languages usually provide the situation to encounter problems in rendering the texts. This matter also subjects translation of ellipsis. For example, in English, all verbs are subject to VPE (example (1), (2), and (3)). In Persian, VPE is possible in presence of some modal verbs (example (2)) or a light verb construction (example (3)), or when the verb is in simple past/present

[1] The form of lexical verb also changes.

passive voice (example (4)). Consider the following examples:

(1) *He usually talks all the time. He didn´t* [*talk all the time*] *yesterday.*

HT
اون معمولا یریز حرف میزنه. دیروز هیچ حرف نزد .
Oon mamoolan yeriz harf mizane. Dirooz hich harfi nazad.
HE/NOM USUALLY CONSTANTLY WORD/NC+HIT/LV/PRS-3SG.
YESTERDAY NOTHING WORD/NC+NEG+HIT/LV/PST-3SG
He usually hits_word constantly.Yesterday not-hit_word nothing.
He usually talks constantly. Yesterday he didn´t talk.

(2) *You cannot live alone forever.*
- *I can [live alone]*.

HT
تو نمیتونـي تا ابد تنهایي زندگي کنـي.
To nemitooni ta abad tanhai zendegi koni.
YOU/NOM-2SG NEG+CAN-2SG TILL EVER ALONE LIFE/NC+DO/LV-2SG
You not-can till ever alone do-you_life.
You cannot live alone for ever.

-من میتونم [تنها زندگی کنم].
Man mitoonam [tanha endegi konam.]
I/NOM CAN/MOD-1SG [ALONE LIFE/NC+DO/LV-1SG]
I can [*alone do-I_life*].
I can [*live alone*].

(3) *I am supporting you all.*
- *Well, don´t* [*support us*].

HT
دارم همتونو تامین میکنم.
Daram hamatoono taaminl mikonam.
HAVE/PRG-1SG ALL+YOU/ACC-2PL SUPPORT/NC+DO/PRS-1SG
Have-I all you do-I_ support.
I am supporting you all.

-خوب [تامین]نکن.
Khob [taamin] nakon.
WELL[SUPPORT/NC+] +NEG+DO/LV/IMP-2SG
Well not- do [*support*].
Well don´t do [*support*].

(4) *Are you bored?*
- *No, I´m not.*

HT
خسته شدی؟
Khaste shodi?
BORED/PP+GET/PPST-2SG
Got-you_bored?
You got bored?

نه [خسته] نشدم.
Na khaste nashodam.
NEG BORED/PP+NEG+GET/PPST-1SG
No, not-got-I_bored.
No, I didn´t get bored.

As noted in the above 4 examples, all the English sentences contain verbal ellipsis; however, the occurrence of VPE is possible in Persian for example (2) and example (4), because, in former, the modal verb میتونم (mitoonam) [CAN/MOD/PRS-1SG] `can-I´ (`I can´) takes the function of the verb phrase تنها زندگی‌کنم (tanha zendegi konam) [ALONE LIFE/NC_DO/LV/PRS-1SG] `do-I_life-I alone´ (`live alone´) and leads into ellipsis, and in later the passive operator شدن(shodan) `get´ takes the function of the verb. In example (3), the antecedent verbal group تامین کردن (Taamin kardan) [ SUPPORT/NC+DO/LV] `do_support´ (`to support´) is a light verb construction consisting of two elements: the nominal component تامین (taamin) `support´ and the light verb کردن (kardan) `to do´. Here, omitting the nominal component and keeping the light verb leads into an ellipsis. This ellipsis is called "NCE", hereafter.

In the following section, the goal is to see how English VPEs adjacent to the auxiliary verbs are dealt with by human translators (HT) and Google Translator (GT), in the Persian language.

4.2.1. HT and GT of VPE in after Auxiliary `Do´

In English, when a verb phrase is omitted and leads into a VPE (example (1) and (3)), or when it is reduced and leads into a so-called pro-verb (example (5)), an auxiliary verb takes the function of the verb. In simple tense, this auxiliary is the verb/pro-verb `do´[2].

[2] `Do´ with its all inflected forms: *do*, *does*, *did*, *done*, and *doing*.

(5) *He never comes.*
- *This time he does.* (`does´ is the pro-verb replacing `comes´)

In this study, 1,025 cases of VPE after auxiliary/pro-verb `do´ were analyzed (Table 2). It was observed that if the verbal group in Persian consists of one lexical element, the human translator fills the resultant gap of ellipsis with the antecedent lexical verb (example (6)) or by its Persian pro-verb counterpart. Persian pro-verb is a combination of اینکارو (inkaro) [THIS+WORK/NC-ACC] `this work´ or همینکارو (haminkaro) [SAME+WORK/NC-ACC] `same-work´ (`the same work´)) and کردن(kardan) `do´ (example (7)).

(6) *He never comes.*
- *This time he does.*

HT
اون هرگز نمیاد.
Oon hargez nemiad.
HE/NOM NEVER NEG+COME/PRS-3SG
He never comes.

اینبار میاد.-
Inbar miad.
THIS/DET+TIME COME/PRS-3SG
This-time comes.
He comes this time.

(7) *He spoiled everything.*
- *He always does.*

HT
او همه چیو خراب کرد.
Oo hamechio kharab kard.
HE/NOM ALLTHING/ACC SPOILT/NC+DO/LV/PST-3SG
He did_spoilt all thing.
He spoiled everything.

همیشه همینکارو میکنه.
Hamishe haminkaro mikone.
ALWAYS SAME+(WORK/N/ACC DO/PR/PRS-3SG)/PV
Always does same-work.
He always does that.

If the verb, in Persian, is a light verb construction, there are three options for the

translator: 1) the gap is filled with the antecedent light verb construction (example (8): "*I don´t* [*need a gun*.]"); 2) the gap is filled with the pro-verb (example (7)); 3) NCE occurs (example (8): "I do.").

(8) *Why you need a gun?*
- *I don´t* [*need a gun*], *my dad thinks I do*.

HT
چرا تو یه تفنگ احتیاج داري ؟
Chera to ye tofang ehtij dari?
WHY/INT YOU/NOM-2S A/DET GUN NEED/NC+HAVE/LV/PRS-2SG
Why you have-you_ need a gun?
Why you need a gun?

- من احتیاج ندارم، پدرم فکرمیکنه که [احتیاج] دارم.
Man ehyiaj nadaram, pedaram fekr mikone ke [ehtiaj] daram.
I/NOM NEED/NC+NEG+HAVE/LV/PRS-1SG FATHER+ MY/POS THOUGHT/NC+DO /LV/PRS- 3SG THAT/CNG [NEED/NC+] HAVE/LV/PRS-1SG
I not-have-I need, my father does-thought that have-I [*need*].
I don´t need, my father thinks that I do.

Another way of translating sentences which confirm a previous statement and involve VPE or pro-verb, is to state همینطوره (hamintore) [LIKEWIES/ADV+BE/PRS-3SG] `likewise is´ (`it is likewise´), as in example (9):

(9) *I thought you hated Kelso.*
- *I do.*

HT
فکر میکردم از کلسو متنفري.
Fekr mikardam az kelso motenaferi.
THOUGHT/NC+DO/LV/PST-1SG FROM KELSO HATE-2SG
Did-I_thought hate-you from Kelso.
I thought you hate Kelso.

- همینطوره.
Hamintore.
LIKEWISE/ADV+BE-3SG
Is likewise.
It is likewise.

They can also be stated with البته (albate) `of course´, حتما (hatman) `certainly´ (hereafter,

they are called “confirming adverbs”) with or without بله (bale) `yes´ preceding them.

Besides the above mentioned patterns of dealing with sentences with VPE and pro-verbs `do´, it was observed that some instances have been translated in a different way, rather than having been translated literally; that is, they have not preserved fidelity, for the sake of naturalness or irony (example (10)). These cases are called, hereafter, “non-literal”.

(10) *I find him.*
- *You do*?

HT
من پیداش میکنم.
Man peydash mikoanm.
I/NOM FINDING/NC+HIM/ACC+DO/LV/PRS-1SG.
I do-I_finding-him.
I find him.

- هیچ کس هم نه تو.
Hichkas ham na to.
NOBODY TOO/PRO NO YOU/NOM-2SG
No nobody too you.
Nobody else, just you.

In the above example, if the part with pro-verb `do´ (i.e. *You do?*) had been translated as a question (literal translation), with keeping the ellipsis as *تو ؟* (to) `you?´ or with substituting a pro-verb as *تو اینکارو میکردی ؟* (To in karo mikardi?) [YOU-2SG THIS/DET WORK/NC+DO/PRS-2SG] `*You do-you this work?´* (*You do that?*), the ironic sense of the sentence would have been damaged and another sense would have been conveyed.

For VPEs or pro-verbs in structures like the comparatives with `than/as…as/like´ or adverbial clauses with `after/before´, another form of translation was observed. English comparative or adverbial clauses are changed (reduction) into Persian comparative or adverbial phrases, respectively (examples (11), (12), and (13)).

(11) *He gives you more courage than I do.*

HT

او بیش از من به تو جرات میده.

Oo bish az man be to jorat mide.

HE/NOM MORE THAN I/NOM TO/PRE YOU/NOM-2SG COURAGE/NC+GIVE/KV/PRS-3SG

He gives_courage to you more than I.

He gives you more courage than I do.

(12) *I know him as long as you do.*

HT

من اونو به اندازه تو میشناسم.

Manam oono be andazeye to mishnasam.

I/NOM HIM-ACC AS+MUCH+AS YOU/NOM-2SG KNOW/PRS-1SG

I know-I him as much as you.

I know him as much as you do.

(13) *We should not move before they do*.

HT

ما نباید قبل از اونا هیچ اقدامی کنیم.

Ma nabayad ghabl az oona hich eghdami konim.

WE/NOM NEG+MUST/MOD BEFORE-OF THEM NOTHING MOVEMENT/NC+DO/LV/PRS-1PL

We not-must do-we_movement nothing before of them.

We must not do anything before they do.

GT usually produces noise in rendering English VPE into Persian when it faces the auxiliary or the pro-verb `do´. `Do´ is treated by GT as a light verb construction signifying *انجام دادن* (anjam dadan) [DO/NC+GIVE/LV] `give_do´ (`do´) and it appears in 3rd person singular regardless of the number and person of the subject[3] (examples (14)). The problem of non-agreement between subject and verb, hereafter, is not taken into consideration.

(14) *You want to kill me?*

- *He does.*

HT

تو میخواي منو بکشی؟

To mikhai mano bokoshi?

YOU/NOM-2SG WANT/PRS-2SG ME/ACC KILL/INF-2SG

You want-you kill-you me?

You want to kill me?

- اون میخواد [منو بکشه].

[3] It should be noted that Persian verbs conjugate differently to show agreement with the subject.

Oon mikhad [mano bokoshe].
HE/NOM WANT/PRS-3SG ME/ACC KILL/INF
He wants [*kill-he me*].
He wants to [*kill me*].

GT
*شما می خواهید مرا بکشند؟[4]**
Shoma mikhahid mara bokoshand?
YOU/NOM-2SG(H) WANT/PRS-2SG(H) ME/ACC KILL/INF-3PL
You want-you kill-they me?
You want they kill me?

*- او انجام می دهد.**
Oo anjam midahad.
HE/NOM DO/NC+GIVE/LVPRS-3SG
He gives_do.
He performs.

Besides, `do´ is always translated as an infinitive verb (indicating no tense): *انجام دهد* (anjam dahad) [DO/NC+GIVE/LV/INF-3SG] `give_do´ (`do´). However, the auxiliaries `does´ and `did´ are translated according to the tense (present and past, respectively); `does´ is translated in *انجام می دهد* (anjam midahad) [DO/NC+GIVE/LV/PRS-3SG] `gives_do´ (`does´), and `did´ in *انجام داد* (anjam dad) [DO/NC+GIVE/LV/PST-3SG] `gave-do ´ (`did´). Therefore, GT fails in recovering VPE after auxiliary `do´ and pro-verbs, due to a) not filling the gap, resultant of ellipsis, with the antecedent verb; b) translating the auxiliary verbs as lexical element; and c) non-agreement in person and number between subject and verb.

It should be noted that the (few) human or Google mistranslation or non-translation of those parts which are not related to ellipsis are not discussed here. It might be claimed that it is possible that bad translation of some parts of the sentence affects the translation of the other parts; consequently, the translation of the entire sentence should be taken into account. However, the scope of this study does not allow covering all these problems. Only minor changes were made in the text, before GT application: a limited preprocessing and normalization of the deficient English sentences by correcting the minor errors regarding punctuations, contracted forms, and some misspelled words, as it was mentioned in chapter 3.

[4] The sign * indicates unacceptable translation.

4.2.2. HT and GT of VPE after Auxiliary `Be´

In this study, 1,054 cases of VPE after auxiliary `be´ were analyzed (Table 1). In English, the auxiliary `be´ (`am, is, are, was, were´) forms part of the present (`am, is, are´) or past (`was, were´) progressive tense as in `she is running´ and also of the simple present or past passive voice as in `the house was built´. This auxiliary can take the function of a verb when the verb drops and leads into a VPE, as in example (15).

(15) *Why our friends were not trying to destroy each other like we were* [*trying to* (*destroy each other*)].

In English, when VPE occurs, the lexical verb, which is in gerund form in the progressive tense and in past participle form in the passive voice, is missing and what remains is the verb `be´.

In Persian, progressive tense is formed by the light verb داشتن (dashtan) `have´ *(`being`)* precedes the lexical verb. In Persian, verbs with progressive tense do not occur elliptically, unless the ellipsis of nominal component in light verb construction (NCE). Therefore, the gap is usually filled by the antecedent verb (example (16)) or by the pro-verb (example (17)).

(16) *Come on Megan*.
- *I am* [*coming*].

HT
بیا مگان.
Bia megan.
COME/IMP-2SG MEGAN
Come Megan.

- دارم میام.
Daram miam.
HAVE/PRSPR-1SG COME/PRS-1SG
Have-I come-I.
I am coming.

(17) *We´re not leaving you.*
- *Yes, you are* [*leaving me*].

HT
ما تنهات نمیزاریم.
Ma tanhat nemizarim.
WE/NOM ALONE+YOU/ACC NEG+LEAVE/PRS-1PL
We not leave-we alone-you.
We don´t leave you alone.

- دارین همینکارو میکنین.
Darin haminkaro mikonin.
HAVE/PRS-2PL+THIS/DET+WORK/ACC+DO/PRSPR-2PL
You have-you do this work.
You are doing so.

In the case of light verb construction, besides two options of filling the gap with antecedent or the pro-verb, NCE can be another form of translating this type of VPE (example (18)).

(18) *Go on driving*.
- *I am* [*going on driving*].

HT
تو رانندگیتو کن.
To ranandegito kon.
YOU/NOM-2SG DRIVING/NC+YOUR/ACC-2SG DO/IMP-2SG
You, do your driving.

- دارم [رانندگی] میکنم.
Daram [ranandegimo] mikonam.
HAVE/LV/PRSPR-1SG [DRIVING/NC]+DO/LV/PRS-1SG
Have-I do-I [*driving*].
I am doing [*drive*].

In Persian, verbs of simple past/present passive voice can occur elliptically. The passive voice, in Persian, is expressed by the presence of passive auxiliary verb شدن (shodan) `get/become´ and the past participle of the main lexical verb as in علي دیده شد (Ali dide shod) [Ali SEE/PP+GET/PPST-3SG] `Ali got seen´ (`Ali was seen´). Some light verb constructions are changed into passive voice by transforming the light verb into passive auxiliary verb شدن (shodan) `get/become´, as example (19).

(19) *Have you convinced him?*
- *He wasn´t* [*convinced*].

HT
متقاعدش کردی؟
Moteghaedesh kardi?
CONVINCED/NC+HIM/ACC+DO/LV//PST-2SG
Did-you Convinced him?
Did you convince him?

-[متقاعد] نشد.
[Moteghaed]] nashod.
CONVINCED/NC+NEG+GET/PPST-3SG
Not got [convinced].
I *He did not get* [*convinced*].

GT**,** in treating with VPE after auxiliary `be´, makes the same noises as in treating with VPE after auxiliary `do´; `be´ is translated like a copula as هستن (hastan) [BE/PRS] `am, is, are´ in present tense (example (20)), and بودن (boodan) [BE/PST] `was, were´ in past tense (examples (21)). `Be´ as a tense or voice operator forms a part of the tense of the verbal group; hence, it should not be translated alone individually.

(20) *Get up Stavan.*
- *I am* [*getting up*].

HT
استوان پاشو.
Stavan pasho.
STAVAN GET/IMP-2SG + UP/PR
Stavan! get up.

- دارم[پا] میشم.
Daram pa misham.
HAVE/PRSPR-1SG+UP/NC+GET/LV/PRS-1SG
Have-I get-I_up.
I am getting up.

GT
بلند شو Stavan .
Boland sho Stavan.
GET/LV/IMP-2SG+UP/NC STAVAN
Get up, Stavan.

- من هستم. *
Man hastam.
I/NOM BE/PRS-1SG
I am-I.
I am.

The HT of the above example contains NCE (ellipsis of پا (pa) [UP/NC] `up´), and the tense, like its English counterpart, is in the present progressive. However, in GT, the auxiliary `be´ has been translated into هستم (hastam) [BE/PRS-1SG] `am´, which has led into a nonsense answer to its previous command.

As mentioned above `be´ in the past tense is translated by GT as بودن (boodan) [BE/PST] `was, were´:

(21) *Maggy was deceived like I was* [*deceived*].

HT
مگی هم مثل من فریب خورد.
Maggi ham mesle man farib khord.
MAGGY ALSO LIKE/COM ME/NOM DECEIT/NC+EAT/LV/PST-3SG
Maggy also ate deceit like me.
Maggy also was deceived like I was.

GT
Maggy مثل من بود فریب بود. *
Magi mesle man bood farib khord.
MAGGY LIKE/COM ME/NOM BE/PST-3SG DECEIT BE/PST-3SG
Maggy was like me was deceit.
Maggy was like me she was a deceit.

4.2.3. HT and GT of VPE after Auxiliary `Have´

In English the auxiliary `have, has, had´ followed by the past participle of the verb, forms the present or past perfect tense. When VPE occurs in the verbal group with perfect tense, this auxiliary remains and the main verb is deleted.

In the present study, 480 cases of VPE after auxiliary ´have´ were spotted and analyzed (table 2). VPE does not occur in Persian with present or past perfect verb; therefore, the gap is recovered through the antecedent verb (example (22)) or the pro-verb (example 23). In case of the Persian light verb construction, NCE is also possible.

(22) *You haven´t eaten your lunch yet?*
- *I have* [*eaten it (my lunch)*].

HT
هنوز ناهار نخوردی[5]؟
Hanooz nahari nakhordi?
YET LUNCH NEG+EAT/PST-2SG
Not-ate-you lunch yet?
You haven´t eaten your lunch yet.

- [ناهار] خوردم.
[Nahar] khordam.
[LUNCH] EAT/PST-1SG
Ate-I [lunch].
I ate [the lunch].

(23) *I trusted on you. I wish I hadn't* [*trusted on you*].

HT
من بهت اعتماد کردم. کاش اینکارو نکرده بودم[6].
Man behet etemad kardam. Ey aksh inkaro nakarde boodam.
I/NOM TO/PRP+YOU TRUST/NC+DO/LV/PST-1SG WISH THIS/DET WORK/ACC NEG+DO/PP+BE/PSTP-1SG
I did-I_trust to-you. Wish not-done_this work_was.
I trusted on you. I wish I hadn´t do that.

GT, in dealing with VPEs in present perfect tense, always gives the output داشته باشد (dashte bashad) [HAVE/PP+BE/INF-3SG] `had-be´ (`have´); that is, it translates the lexical verb `have´ signifying داشتن (dashtan) `possess´ in conditional mood (example (24)).

(24) *I love you, I always have* [*loved you*].

GT
من شما را دوست دارم ، من همیشه داشته باشد.*
Man shoma ra doost daram, man hamishe dashte bashad.
I/NOM YOU/ACC(H) LOVE/NC_HAVE/LV/PRS-1SG I/NOM ALWAYS HAVE/PP+BE/INF-3SG
I have-I_love you, I always had-be.
I love you, I always has.

Google treats all the combination of "subject+ `already has/have.´" as "subject + در حال حاضر (dar hale hazer) `at present time´", as in example (25).

(25) *You´re not going to talk with him.*

[5] In colloquial texts (dialogues), Persian present perfect tense appears in form of simple past tense. Otherwise it is formed by past participle of lexical verb + the tense operator in form of a prefix attached to the main verb (conjugated according to the person and number of subject).
[6] Past perfect tense, in Persian, is formed by past participle form of lexical verb preceded by auxiliary verb بودن (boodan) [BE/PSTP] `was/were´ which is conjugated according to the person and number of subject.

- *I already have* [*talked with him*].

HT
نمیخوای باهاش حرف بزنی؟
Nemikhahibahash harf bezani?
NEG+WANT7PRS-2SG WITH+HIM/NOM WORD/NC+HIT/INF-2SG
Not-want-you hit-you_ word with-him?
You don´t want to speak with him?

-[حرف] زدم.
harf zadam.
WORD/NC+HIT/PST-1SG
Hit-I_word.
I did.

GT
*شما در حال رفتن به دور.**
Shoma dar hal raftan be door.
YOU/NOM-2SG(H) BEING+GO/GR TO/PRP FAR
You being-going to far.
You in being going far.

*- من در حال حاضر. *
Man dar hal hazer.
I/NOM AT/PRO-PRESENT
I at present.
I at present time.

All VPEs in the past perfect tense are translated as "subject + *تا به حال* (ta be hal) [TILL+TO/PRE/NOW] `till to now´ (`up to now´).

To sum up, VPE after auxiliaries `do´, `be´, and `have´ (2,559 cases) are translated by human translators as follows:

1. If the English verbal group is not a light verb construction in Persian (which happened in 22.24% of the VPE here studied), the VPE is treated as one of the followings:

a) The gap produced by VPE is filled by its antecedent (13.46%): (Subject) + (negation element) + *antecedent verb* (in tense matching with antecedent verb), suffixed by non-emphatic subject and matching with emphatic subject in number and person.

b) (Subject) + (negation element) + *pro-verb* (in tense matching with antecedent verb) suffixed by non-emphatic subject which matches with emphatic subject in number and person (8.78%).

2. If the English verbal group is a light verb construction in Persian (45.37%), besides the above two forms of translation (a) and (b), there is a third form (c) in which the Persian sentence carries an ellipsis but of a different kind, distinct from that of English:

a) (Subject) + *antecedent verb* (9.66%); the light verb is suffixed by non-emphatic subject and match with emphatic subject in number and person. In case of negation, the light verb is prefixed to the negation element.
b) Subject) + (negation element) + *pro-verb* (in tense matching with antecedent verb) suffixed by non-emphatic subject which matches with emphatic subject in number and person (8.49%).
c) (Subject) + (negation element) + *the light verb* (27.22%) (in tense matching with antecedent verb, suffixed by non-emphatic subject and match with emphatic subject in number and person).

3. For VPE passive voice after auxiliary `be´ the ellipsis can be retained (1.13%).

4. Or in case of confirming a previous statement, the translation is بله (bale) `yes´ with/without a confirming adverb following it (23.02%).

5. With English VPE in comparative and adverbial clauses, the clauses change into Persian adverbial or comparative phrases (4.30%).

6. *Non-literal translation* forms (2.38%).

Table 2a and 2b, and also Fig 1 sum up these observations. Table 2a continues in Table 2b, which presents the other possible translation strategies applicable for both lexical verbs and light verb constructions.

Table 2a. HT of VPE after operators (`do, be, have´)

Operators	Cases	Lexical verbs			Light verb constructions			
		VPE	Antecedent	Pro-verb	VPE	Antecedent	Pro-verb	NCE
do	1,025	–	13.46%	8.78%	–	9.66%	8.49%	27.22%
be	1,054	0.85%	12.15%	8.54%	0.28%	5.98%	3.98%	18.60%
have	480	–	36.87%	2.92%	–	16.88%	5.43%	19.18%
Total	2,559	0.35%	17.31%	7.58%	0.12%	9.50%	6.06%	22.16%

Table 2b. HT of VPE after operators (`do, be, have´)

Lexical verbs/light verb constructions				
Operators	Confirming statements	Comparatives/Adverbials	Non-literals	Wrong/No Translation
do	23.02%	5.27%	2.83%	1.27%
be	41.75%	4.65%	2.75%	0.47%
have	16.63%	1.47%	0.62%	–
Total	29.54%	4.30%	2.38%	0.70%

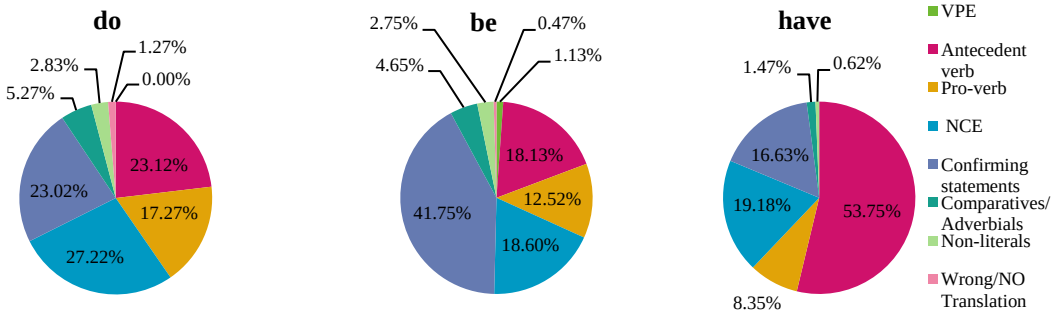

Figure 1. HT of VPE after operators (`do, be, have´)

As the above tables and figure shows the HT tendency in translation of VPE after the operators is mostly toward keeping the original structure (table 2a: 63.08%) than following the other structures (table 2b: 36.92%).

VPE after auxiliary `have´ is mostly recovered in Persian by the antecedent verb (53.75). It seems that the majority instances of VPE after auxiliary `be´ were in linguistic construction of a confirming answer to a previous statement; for which the adopted translation strategy was using the confirming statement (41.75%). The strategies for other instances of VPE after this auxiliary were quite similarly divided among the strategies of the antecedent verb (18.13%), pro-verb (12.52%), and NCE (18.60%). It is worth mentioning that NCE is, in fact, a sub-category of VPE, as the light verb component is deleted and the nominal component is retained. VPE after auxiliary `do´

were translated using quite the same number of occurrences of strategies: antecedent verb (23.12%), pro-verb (17.27%), NCE (27.22%), and confirming statement (23.02%). A residual number of cases were translated using different constructions (comparative/adverbials) or were translated non-literally.

4.2.4. HT and GT of VPE after Modal Verbs

In English, a modal verb can take the function of a lexical verb allows for a VPE, whereas in Persian, in fact the choice of the modal affects the possibility to form VPE. VPE can occur with modal verbs of ability or possibility (*توانستن* (tavanestan) `can´ and *ممکن بودن* (momken boodan) [POSSIBLE-BE-GR] `being possible´ (`may´), respectively) and with the modal verbs of obligation `must/have to´ if they are translated as *مجبور بودن* (majboor boodan) [OBLIGED+BE/GR] `being obliged´). On the other hand, VPE cannot occur with modal verbs of obligation *باید* (bayad) `should´ and modal verbs of volition *خواستن* (khastan) `will´[7].

In this study, 1,135 instances of VPE after modal verbs were analyzed. In the following, it is aimed to describe these models individually and to assess how GT treats with VPE in presence of each modal verb:

4.2.4.1 VPE after the Modal Verbs `Can´ and `May´

English VPE after modals `can´ and `may ´ can be treated in Persian either by recovering the gap by the antecedent verb or by keeping the same ellipsis (example (26)).

(26) *Laugh while you can* [*laugh*].

HT
بخند تا میتونی[8] [بخندی] .
Bekhand ta mitooni [bekhandi].
LAUGH/IMP-2SG TILL CAN/MOD/PRS-2SG+[LAUGH/INF-2SG]
Laugh till can-you[*_laugh-you*].

[7] It is also a tense operator; it is mentioned in modal group due to its similarity with the other modal verbs.
[8] In Persian, both lexical verbs and modal verbs `can/could´ and `will/would´, in a verbal group, are conjugated by the subject pronoun, which indicates the number and person of the verb.

Laugh while you can [*laugh*].

In case of the Persian light verb constructions, besides these above mentioned processes, the nominal component can be omitted. In case of confirming statements, a confirming adverb, with or without *بله* (bale) `yes´ is the output translation.

GT retains the ellipsis after the modal verb *توانستن* (tavanestan) `can´ and produces an acceptable output, as in the example below which is the translation of example (26):

GT
بخند در حالیکه شما می توانید [بخندید].
Bekhand dar halike shoma mitavanid [bekhandid].
LAUGH/IMP WHILE YOU/NOM-2SG(H) CAN/MOD/PRS-2SG(H) [LAUGH/INF-2SG(H)]
Laugh while you can-you [*laugh*].
Laugh while you can [*laugh*].

In the above example, the only problem is that the GT has not observed consistency regarding honorific. English-Persian GT usually translates the sentences in honorific form; that is, the 2nd person pronoun (singular) *تو* (to) `you´ is translated as *شما* (shoma) `you´, which is also the plural 2nd person pronoun. Accordingly, the verb also conjugates differently; that is, it is also expressed in plural form as in *می توانید* (mitavanid) [CAN/MOD/PRS-2SG(H)] `can-you´ or *بخندید* (bekhandid) [LAUGH/INF-2SG(H)] `laugh´. In this regard, it is expected that GT translates ´laugh´ as *بخندید* (bekhandid) [LAUGH/INF-2SG(H)] `laugh´ rather than *بخند* (bekhand) [LAUGH/INF-2SG] `laugh´.

GT fails in treating with VPEs after `can´ in conditional sentences or sentences containing verbs of subjunctive mood[9]. In these sentences, the modal verb must be translated in infinitive form; that is, it should be prefixed by the condition morpheme ب (b)[10]). However, it was noticed that GT renders this modal verb in present/future tense with prefix *می* (mi), which leads into an unacceptable output (example (27)).

(27) *I cannot go along with this.*

[9] Verb mood to express a wish, emotion, possibility, judgment, opinion, necessity, or action that has not yet occurred - like *`Before he does, If I can, I hope you come, I don´t think they agree, Perhaps we go ...´*
[10] In Persian, this morpheme forms both imperative and infinitive forms of verb.

- *I hope you can* [*go along with this*].

HT
نمیتونم با این ادامه بدم.
Nemitoonam ba in edame bedam.
I NEG/CAN/MOD/PRS-1SG WITH THIS/PRO CONTINUATION/NC+GIVE/LV/INF-1SG
Not-can-I give-I_continuation with this.
I cannot continue with this.

امیدوارم بتونی[*ادامه بدی*] .
Omidvaram betooni [edame bedi].
HOPE/PRS-1SG CAN/MOD/INF-2SG [CONTINUATION/NC+GIVE/LV/INF-2SG]
Hope-I can-you [*give-you continuation*].
I hope you can [continue].

GT
*من نمی توانم با این.**
Man nemitavanam ba in.
I/NOM NEG+CAN/MOD/PRS-1SG WITH THIS/DET
I not-can-I with this.
I cannot with this.

*- من امیدوارم که شما می توانید.**
Man omidvaram ke shoma mitavanid.
I/NOM HOPE/PRS-1SG THAT/RPRO YOU-2SG(H) CAN/MOD/PRS-2SG(H)
I hope-I that you can-you.
I hope that you can.

GT generates noise when the modal verb *توانستن* (tavanestan) `can´ is in past tense; the tense is not preserved: `Could´ is translated in present/future tense as in example (28).

(28) *I did the best I could* [*do*].

HT
من هر کاری که تونستم [*بکنم*] *کردم.*
Man kar kar ke toonestam [bokonam] kardam.
I/NOM-1SG ANY/DET WORK/NC THAT/RPRO CAN/MOD/PST-1SG [DO/LV/INF-1SG] DO/PST-1SG
I did-I_any work could-I[*do-I*].
I did whatever I could [do].

GT
*من بهترین من می تواند انجام داد.**
Man behtarin man mitavanad anjam dad.
I/NOM BEST I/NOM CAN/MOD/PRS-3SG DO/NC- GIVE/LV/PST-3SG
I best I can give-do.
I did I best can.

As for modal verb of possibility `may´, in Persian, this modal appears in adjective construction form, rather than a verb: ممکن است / ممکنه[11] (momken ast) [PROBABLE/BE/PRS-3SG] `it is probable´ (`may´)). This construction, like English, can lead into the ellipsis; however, it results in ellipsis of subject as well (example (29)) because this modal cannot be conjugated with the subjects. This modal is, in fact, an adjectival construction with sentential subject and a modal value.

(29) *You are leaving him?*
- *I may* [*leave him*].

HT
حالا میخوای ترکش[12] کنی ؟
Hala mikhai tarkesh koni?
NOW WANT/PRS-2SG LEAVE/NC+HIM/ACC+DO/LV/INF-2SG
Now want-you do-you_leave-him?
Now, you want to leave him?

- ممکنه [ترکش کنم].
Momkene [tarkesh konam].
PROBABLE+BE/PRS-3SG [LEAVE/NC+HIM/ACC+DO/LV/INF-1SG]
Is probable[*do-I_leave-him*].
[*I*] *may* [*leave him*].

GT translates VPE after `may´ as "subject + ممکن است (momken ast) [PROBABLE/BE/PRS-3SG] `it is probable´, as in the following example which is the translation of example (29).

GT
شما او را ترک؟ *
Shoamoora tark?
YOU/NOM-2SG(H) HIM/ACC LEAVE/NC
You him leave.

- من ممکن است. *
Man momkene ast.
I/NOM PROBABLE+BE/PRS-3SGI
I is probable.
I may.

[11] ممکنه is the contraction form of ممکن است.

[12] In Persian, an accusative pronoun can be suffixed (in a bound morpheme form) to the nominal component of a light verb construction, in colloquial texts. However, in other texts, this pronoun follows the verb as a free morpheme, as in GT of the same example: او را (oora) [HIM/ACC] `him´.

In the foregoing example, GT translates the modal verb correctly but because of the differences of modal structure in the two languages, it fails in translation; it keeps the subject and the result is unacceptable.

4.2.4.2. VPE after the Modal Verb `Will´

Persian future tense in informal speaking usually appears in form of simple present tense (i.e., without future operator); therefore, in Persian corpus, the English future tenses with auxiliary `will´ have been translated in simple present tense; however, GT always produces formal utterances (examples (32) and (33)); thus, the modal verb indicating the future tense precedes the main verb.

As mentioned before, VPE after `will´ does not occur in Persian. There is just possibility of NCE for light verb construction. Hence, the gap produced by English VPE, is filled out by the antecedent verb (example (30)) or pro-verb in Persian.

(30) *If I need to go I will* [*go*].

HT
اگر لازم باشه برم میرم.
Agar lazem bashe beram miram.
IF NECESSARY BE/INF-3SG GO/INF-1SG GO/PRS-1SG
If is necessary go-I go-I.
If it is necessary that I go, I will go.

In Persian light verb constructions, NCE can also occur (example (31)).

(31) *If we wanted to harm you we would* [*harm you*].

HT
اگه میخواستیم آزارت بدیم که [آزارت] میدادیم.
Age mikhastim azaret bedim ke [azaret] midadim.
IF WANT/CON/PST-1PL HARM/NC+YOU/ACC-2SG+GIVE/LV/INF-1PL THAT/CNJ [HARM/ACC+YOU/ACC-2SG]+GIVE/LV/CON/PST-1PL
If wanted-we give-we_harm-you that give-you[*_harm-you*].
If we wanted to harm you, then we would [*harm you*].

GT translates all VPEs with `will´, from English into Persian, as "subject + خواهد (khahad) [WILL/MOD-3SG] `will´ + شد (shod) [GET/P-3SG] `get´, as in example (32).

(32) *If I need to learn I will* [*learn*].

GT
*اگر من باید یاد بگیرند من خواهد شد.**
Agar man bayad yad begiranad man khahad shod.
IF I/NOM MUST/MOD LEARNING/NC+TAKE/LV/INF-3PL I/NOM WILL/MOD-3SG GET/PST
If I must take_learning-they I will-s/he get.
If I must learn I will be [*learnt*].

VPE after `would´ is translated by GT as "subject + خواهد (khahad) [WILL/MOD-3SG] `will´ + بود (bood) [BE/INF-3SG] `be´.

(33) *If we wanted to harm you we would* [*harm you*].

GT
*اگر ما می خواستیم به شما ضرر ما خواهد بود.**
Agar ma mikhastim be shoma zarar ma khahad bood.
IF WE/NOM WANT/ON/PST-1PL TO/PRE YOU/2SG(h) HARM/NC WE WILL/MOD-3SG BE/INF-3SG
If we wanted-we harm to you we will be.
If we wanted harm you we will be.

4.2.4.3. VPE after the Modal Verbs `Must/Have to/Should´

The English modal verbs of obligation `must´ and `have to´ have two Persian counterparts which have the same meaning but expressed differently: one is باید (bayad) and the other is مجبور بودن (majboor boodan) [OBLIGED+BE/GR] `being obliged´. مجبور بودن (majboor boodan) is conjugated in according to the person and number of subject. `Should´ is also translated as باید (bayad).

Persian modal verb مجبور بودن (majboor boodan), in contrast with باید (bayad), can take the function of the main verb and create VPE. Consider examples (34) and (35), in which `must´, has been translated into Persian in two different ways, in example (34) the Persian translation has not kept the ellipsis; whereas, in example (35) the Persian translation has retained the ellipsis.

(34) *I must go.*

- *Yeah you must* [*go*].

HT
من باید برم.
Man bayad beram.
I/NOM MUST/MOD GO/INF-1SG
I must go-I.
I must go.

- آره باید بري.
Are bayad beri.
YEAH MUST/MOD GO/INF-2SG
Yeah must go-you
Yeah, you must go.

(35) *I´ll try it alone. I have to* [*try it alone*].

HT
خودم تنهایی انجامش بدم. مجبورم [تنهایی انجامش بدم].
Khodam tanhai anjamesh bedam. Majbooram [tanhai anjamesh bedam].
MYSELF/REF ALONE DO/NC+IT/ACC+GIVE/LV/PST-1SG OBLIGED+BE/PRS-1SG [ALONE DO/NC+IT /ACC+ GIVE / LV/PST-1SG]
Myself alone give-I_do-it. I-am_obliged[alone give-I_do-it].
I do it alone. I have to.

Therefore, if the modal `must´ is translated into *باید* (bayad) `must´, the gap is filled by its antecedent verb or a pro-verb; if it is translated into *مجبور بودن* (majboor boodan) [OBLIGED+BE/GR], the lexical verb can be deleted.

GT translates the English the VPEs with modals `must/should´ as "subject + *باید* (bayad) `must´", that is, it translates just two elements which remain after ellipsis occurs (subject and the modal verb). Consider the Google translation of the example (34):

GT
من باید بروم.
Man bayad beram.
I/NOM MUST/MOD GO/INF-1SG
I must go-I.
I must go.

*- آره شما باید.**
Are to bayad.
YEAH YOU/NOM-2SG(H) MUST/MOD
Yeah you must.

GT in dealing with VPE in after modal verb `have to´ produces a nonsense translation. In present tense (`have to/has to´) it is rendered as "subject suffixed to را (ra) (accusative operator) + به (be) [TO/PRE] `to´", as in Google translation of the example (35):

GT

من آن را به تنهایی امتحان کنید. من را به.*

Man anra be tanhai emtehan konid. Man ra be.

I/NOM IT/ACC TO/PRP ALONE TRY/NC+DO/LV/IMP-2SG. I/ACC TO/PROP

I try-you it to alone. Me to.

I, try it alone. Me to.

Although in the above mentioned VPEs after operators and the other modal verbs, the matter of tense did not make any change in the performance of GT, here, GT, for past tense of modal of obligation (`had to´), produces a total irrelevant output as "subject + تا به حال (ta be hal) [TILL+NOW] `up to now´" (example (36))

(36) *I never thought about the past. I never had to.*

HT

من هیچوقت به گذشته فکر نکردم. مجبور نبودم[فکر کنم] .

Man hichvaght be gozashte fekr nakardam. Majboor naboodam [fekr konam].

I/NOM NEVER/ADV TO/PRO PAST/NOM THOUGHT/NC+NEG+DO/LV/PST-1SG. OBLIGED+BE/PST -1SG [THINK/NC+DO /INF-1SG].

I never did-I_thought to past. I was-not_obliged.

I never thought about past. I didn´t have to.

GT

من درمورد گذشته هرگز فکر نمی کردم. من تا به حال هرگز به.*

Man darmorede gozashte hargez fekr nemikardam. Man ta be hal hargez be.

I/NOM ABOUT/PRP PAST/NOM NEVER/ADV THOUGHT/NC+NEG+DO/LV/PST-1SG. I/NOM TILL TO/PRP NOW NEVER/ ADV TO/PRP

I never did-I_thought about past. I till to now never to.

I never thought about past. I never to up to now.

To sum up, HT of VPE in presence of modal verbs (1,135 cases) is resumed as follows:

1. For modals `can/may´ (457 cases), and `must/have to´, if they are translated as بودن مجبور (majboor boodan) [OBLIGED+BE/GR]) (cases143), for lexical verbs:

a) The *VPE* is retained (28.01% and 10.51%, respectively);
b) The gap is recovered by the *antecedent verb* (7.44% and 0.85%, respectively);

c) Or the gap is recovered by the *pro-verb* (2.84% and 6.53%, respectively).

If the English verbal group is a light verb construction in Persian:

a) The *VPE* is retained (21.01% and 17.05%, respectively);
b) The gap is recovered by the *antecedent verb* (20.63% and 5.68%, respectively);
c) The gap is recovered by the *pro-verb* (4.38% and 0%, respectively);
d) Or *NCE* occurs (14.66% and 0%, respectively).

2. For other modals which cannot lead into VPE in Persian (i.e., `will´ (326 cases) and `should/must/have to´, if they are translated into بایَد (bayad) `must´ (cases 166):

a) The gap is recovered by the *antecedent verb* (35.28% and 18.18%, respectively);
b) Or the gap is recovered by the *pro-verb* (4.30% and 3.13%, respectively).

If the English verbal group is a light verb construction in Persian:

a) The gap is recovered by the *antecedent verb* (18.71% and 12.22%, respectively);
b) The gap is recovered by the *pro-verb* (6.75% and 4.55%, respectively);
c) Or *NCE* occurs (16.87% and 9.09%, respectively).

3. *Non-literal translations* (1.23% in total out of 1,135 cases) and *confirming statement*s (6.34% in total out of 1,135 cases) are other possible translation strategies.

4. Other instances of VPE after modal verbs in comparative and adverbial structures, or in confirming statements, show the same pattern as those formed after operators (8.37% in total out of 1,135 cases).

Table 3a and 3b, and also Fig. 2 sum up these observations.

GT in dealing with VPE with modal verb `can´ produces noise in tense and mood of the modal; with `will´ the translation is “subject + modal verb in passive tense”; with `may´

and `must/should´ it gives the unnatural combination of subject followed by the modal verbs; and with `have to´ the output is totally nonsense.

Table 3a. HT of VPE after modal verbs

Modals	Cases	Lexical verbs			Light verb constructions			
		VPE	Antecedent	Pro-verb	VPE	Antecedent	Pro-verb	NCE
Can/may	457	28.00%	7.44%	2.84%	21.01%	2.63%	4.38%	14.66%
Will	326	–	35.28%	4.30%	–	18.71%	6.75%	16.87%
must/have to/should	352	10.51%	19.03%	9.66%	17.05%	17.90%	4.55%	9.09%
Total	1,135	14.54%	19.03%	5.38%	13.75%	11.98%	5.11%	13.57%

Table 3b. HT of VPE after modal verbs

Lexical verbs/light verb constructions				
Modals	Confirming statements	Comparatives/Adverbials	Non-literals	Wrong/No Translation
Can/may	2.63%	15.97%	0.44%	–
Will	7.36%	4.60%	3.68%	2.45%
must/have to/should	10.22%	1.99%	–	–
Total	6.34%	8.37%	1.23%	0.70%

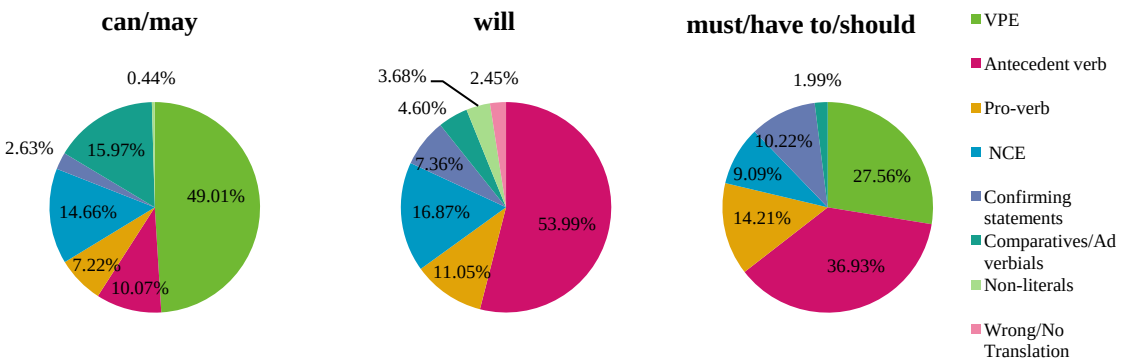

Figure 2. HT of VPE after modal verbs

According to the tables and figure above, here again, the translators have tried to keep the original structure in majority of cases (table 3a: 83.36%).

The structures with `can/may´ were mostly translated into Persian by keeping the VPE (49.02%); the structures with `will´ were mostly translated using the antecedent verb (50.99%); and the structures with `most/have to/should´ kept the VPE for 27.56% of cases and used the antecedent verb in translation for 36.93% of cases.

4.3. VPE after Infinitival Complementizer `to´

In English, VPE occurs frequently after infinitival complementizer `to´. In Persian corpus, this kind of ellipsis (654 cases) has been treated in three ways: a) in the majority (513 cases), the sentence did not contain ellipsis (example (37)); b) the ellipsis was kept just in 127 cases (example (38)); c) and in very insignificant cases (12 instances) the ellipsis was filled by the pro-verb, as in example (39).

(37) *I kept my mouth shut because they wanted to*. (That is ...*they wanted* [*me*] *to* [*keep my mouth shut*]; or …*they wanted* [*me*] *to* [*do so*].)

HT
اونا ازم خواستن که حرفی نزنم.
Oona azam khastan ke harfi nazanam.
THEY/NOM FROM/PRE+ME WANT/PST-3PL THAT/C WORD/NC+ANY/DET+NEG+HIT/LV/INF-1SG
They wanted-they from me that not-hit_any-word.
They wanted me that I do not talk.

(38) *I didn´t want to come, I was forced to* [*come*].

HT
من نمیخواستم بیام، مجبورم کردن [بیام].
Man nemikhastam biam, majbooram kardan [biam].
I/NOM NEG+WANT/PST-1SG COME/INF-1SG FORCE/NC+ME/ACC+DO/LV/PST-3PL [COME/INF-1SG]
I not-wanted-I come-I, did_they-force-me [come-I].
I didn´t want to come, they forced me[to come].

(39) *I offended her, I didn´t intend to* [*offend her*].

HT
اذیتش کردم، نمیخواستم اینکارو کنم.
Azyatesh kardam, nemikhastam inkaro konam.
OFFENCE/NC+HER/ACC+DO/LV/PST-1SG NEG+WANT/PST-1SG THIS/DET+WORK/ACC DO/INF-1SG
Did-I_offence-her, not-wanted-I do this work
I offended her, I didn´t want to do so.

Table 4 summarizes these observations.

Table 4. HT of VPE after complementizer `to´

Number of cases	VPE	Antecedent verb	Pro-verb
657	19.42%	78.44%	2.14%

It was noticed that, in English, the verbs which precede the infinitival complementizer

`to´ do not vary much; namely: `want, tend, expect, try, pretend, need, seem, used, wish, dare, commit, attempt, like (love), and mean´.

GT retains this kind of ellipsis in all cases; hence, the output is often unnatural even if interpretable in most cases (example (40)).

(40) *I kept my mouth shut because they wanted to*.

GT

من دهانم را بسته نگه داشته چراکه آنها می خواستند [دهانم را بسته نگه دارم].

Man dahanam ra baste negah dashte chera ke anha mikhastan [dahanam ra baste negah daram].

I/NOM MOUTH+MY/ACC SHUT/PP KEEP/NC+HAVE/LV/PPER BECAUSE THEY WANT/PST-3PL [MOUTH+ MY/ACC SHUT/ PP KEPT /NC+HAVE/LV/INF-1SG].

I have-had_kept my mouth shut because they wanted-they [*have-I_kept shut*].

I kept my mouth shut because they wanted [*I keep my mouth shut*].

The output would be more natural if, like HT of the same example, the gap after the verb می خواستند [mikhastand] had been filled with the antecedent verb or pro-verb, as it is the most common process in Persian.

4.4. VPE in the Pro-forms `so/as well/also/too/either/neither´

The identical comparatives `so/as well/also/too´ and `neither/either´ are often used to agree with a positive or negative statement, respectively. They are used to avoid the repetition of verbs or verb phrases and the function of verb is taken by auxiliary verbs, as in *John is a teacher and so is his uncle)* and *She does not play guitar and neither do I.* Persian pro-forms are structured as follows:

- "Subject +م[13]/ هم (am/ham) `too´ (+همینطور (hamintor) `likewise´)".
- "Subject + هم/ م (am/ham) `too´+ ((ن) (n) (negative morpheme))[14] + full verb/pro-verb".

They are different just in form, as it is known a sentence can be expressed in different

[13] م (m) `too´ is the contraction form of هم (ham) `too´.

[14] In *neither*/*either* construction.

forms with the same meaning.

From among 175 cases of VPE in pro-form structures, they were mostly translated into Persian pro-form structures; that is, the ellipsis was kept. In a residual number of cases, the gap was filled with the antecedent verb. In the following example (41), the English pro-form has been translated into Persian with filling the gap with the antecedent verb:

(41) *When she died so did I.*

HT
وقتي اون مرد من هم مردم.
Vaghti oon mord mna ham mordam.
WHEN/RPRO SHE DIE/PST-3SG I/NOM ALSO DIE/PST-1SG
When she died I too died-I.
When she died I died too.

The 175 English pro-forms, except 9 cases, (regardless of the tense or the kind of auxiliary verbs), have been translated by human translators into the Persian pro-form structures. The rest 9 cases included the antecedent verb as well, as in example (42).

(42) *They didn´t choose it neither did I.*

HT
اونا انتخابش نکردن منم انتخاب نکردم .
Oona entekhabesh nakardan manam entekhab nakardam.
THEY/NOM CHOICE/NC+IT/ACC+NEG+DO/LV/PST-3PL I/NOM+ TOO/PF CHOOSE/NC+NEG+DO / LV/ PST-1SG
They not-did-they_ choice-it_ I also not-did-I _choose.
They did not choose it neither did I.

In the following, the way Google translator renders these structures is discussed below. One proper performance of GT observed in all pro-forms is that it diversifies the site of the auxiliary verbs and the subjects to produce an order of a statement; otherwise the sentence takes an interrogative structure.

4.4.1 GT of VPE in the Pro-form `so´

Google rendered the pro-forms `so´ in two different ways. From among 67 cases, 30 of

which were translated by GT into Persian pro-forms (12 instances translated by GT properly (example (43)) but 18 instances of them contained problems of translating auxiliary verbs and non-agreement between subject and verb in person and number (example (44)); and 37 instances (about half) were translated as adverbial clauses in which `so´ played the role of a conjunctive adverb signifying *بنابراین* (banabrin) `therefore´, along with the aforementioned problems of translation and non-agreement (example (45)).

(43) *I pity the French.*
- *So do I.*

HT
به فرانسوییا رحمم میياد.
Be faransaviha rahmam miyad.
TO/PRE FRENCH/PL MERCY/NC+MY COME/LV/PRS-1SG
Come_my-mercy to Frenches.
I pity the French.

- من هم همینطور.
Man ham hamnitor.
I/NOM TOO/PF LIKEWISE/ADV
I too likewise.
I do too.

GT
*من ترحم فرانسه است.**
Man tarahom faranse hast.
I/NOM MERCY FRANCE BE/PRS/3SG
I is France mercy.

- من هم همین طور.
Man ham hamnitor.
I/NOM TOO/PF LIKEWISE/ADV
I too likewise.
I do too.

(44) *When she died so did I.*

GT
*هنگامیکه او درگذشت من هم انجام داد.**
Hengami ke oo dargozasht man ham anjam dada.
WHEN/PPRO HE/NOM DIE/PST-3SG I/NOM TOO/PR DO/NC+GIVE/LV/PST-3SG
When he died I gave-do too.
When he died I did too.

(45) *The sun goes to bed and so do we.*

HT

خورشید می ره بخوابه ما هم همین طور.

Khorshid mire bekhabe ma ham hamintor.

SUN GO/PRS-3SG SLEEP/INF-3SG WE/NOM TOO LIKEWISE/ADV

Sun goes sleep we too likewise.

The sun goes to sleep we do too.

GT

خورشید به بستر می رود و بنابراین ما انجام دهد.*

Khorshid be bastar miravad banabarin ma anjam dahad.

SUN GO/PRS-3SG TO/PRE BED AND THEREFORE WE/NOM DO/INF-3SG

Sun goes to bed and therefore we do.

4.4.2. GT and VPE in the Pro-form `as well/also´

23 instances of pro-forms `as well´ and `also´ were observed; they were translated (by GT) into Persian pro-form structure but containing the problem(s) of none-agreement between subject and verb, translating the auxiliary verb, and/or not observing the tense (example (46)).

(46) *I could forgive him.*
- *I could* [*forgive him*] *also.*

HT

میتونستم ببخشمش.

Mitoonestam bebakhshamesh.

CAN/MOD/PST-1SG FORGIVE/INF-1SG+HIM/ACC

Could-I forgive-I-him.

I could forgive him.

- منم [میتونستم ببخشمش] .

Manam [mitoonestam bebakhshamesh].

I/NOM+ALSO/PR [CAN/MOD/PST-1SG FORGIVE/INF-1SG+HIM/ACC].

I also [*could-I forgive-I-him*].

I[*could*] *also* [*forgive him*].

GT

من می توانم اورا ببخشند.*

Man mitavanam oora bebakhsham.

I/NOM CAN/MOD/PRS-1SG HIM/ACC FORGIVE/INF-3PL

I can-I forgive-they him.

I can forgive him.

- من نیز می توانند [اورا ببخشم].
Man niz mitavananad oora bebakhshand.
I/NOM ALSO CAN/MOD/PRS-3PL [HIM/ACC FORGIVE/INF-1SG].
I can-I also [*forgive-I him*].
I can also [*forgive him*].

The above GT is acceptable; except for `could´ rendered in present tense.

4.4.3. GT and VPE in the Pro-form `too´

20 VPE instances were found in Pro-forms with `too´. Two different translations were observed: a) in 14 cases, the construction has been translated as a pro-form in Persian (example (47)); b) in the rest (6 cases) `too´ has been taken as a degree adverb signifying بیش از حد (bish az had) [MORE/COM THAN ENOUGH] `excessively´ (example (48)).

(47) *I forgot it. You can* [*forget it*] *too.*

HT
من فراموشش کردم. تو هم میتونی [فراموشش کنی] .
Man faramooshesh kardam. To ham mitooni [faramooshesh koni].
I/NOM FORGET/NC+IT/ACC+DO/LV/PST-1SG YOU/NOM-2SG TOO/PF CAN/MOD /PRS-2SG [FORGET/NC+IT/ACC+DO / LV/ INF -2SG]
I Did-I_forget-it. You too can-you [*do-you_forget-it*].
I forgot it. You can [*forget it*] *too.*
GT
من آن را فراموش کرد.* شما هم می توانید [فراموشش کنید] .
Man anra faramoosh kard. Shoma ham mitavanid [faramoosh konid].
I/NOM IT/ACC FORGET/NC+DO/LV/PST-3SG. YOU/NOM-2SG(F) TOO/PF CAN/ MOD /PRS-2SG(F) FORGET/NC+DO/LV/INF-2SG(F)
I did_forget it. You can [*do_forget it*] *too.*
I forgot it. You can [*forget it*] *too.*

(48) *Many people have died for this crown, he will* [*die for this crown*] *too.*

HT
عده زیادی به خاطر این تاج مردن اونم می میره.
Eddeye ziadi bekhater in taj mordan oonam mimire.
NUMBER MANY FOR/PRE THIS/DET CROWN DIE/PST-3PL HE+TOO/PF DIE/PRS-3SG.
Many number died-they for this crown he dies-he too.
Many people have died for this crow he dies too.

GT
بسیاری از مردم برای این تاج درگذشت، او بیش از حد.*
Besyari az mardom baraye taj dargozasht, oo bish az had.
MANY OF PEOPLE FOR/PRE THIS/DET CROWN DIE/PST-3SG, HE/NOM MORE/COM THAN ENOUGH
Many of people died for this crown, he excessively.

4.4.4. GT and VPE in Pro-form `Either´

10 instances of VPE were observed in pro-form structures with `either´. GT translated them totally nonsense as "subject + یا(ya) `or´ + نه(na) `no´", as in example (49).

(49) *I do not know this.*
- *I do not* [*know this*] *either.*

HT
من اینو نمیدونم.
Man balad nistam.
I/NOM THIS/ACC NEG+KNOW/PRS-1SG
I not-know-I this.
I don´t know this.

- من هم همینطور[15].
Man ham hamintor.
I/NOM TOO/PF LIKEWISE/ADV
I too likewise.
I do not either.

GT
من نمی دانم درمورد این.
Man nemidoonam darmorede in.
I/NOM NEG+KNOW/PRS-1SG ABOUT/PRE THIS
I not-know-I about this.
I don´t know about this.

- من یا نه.*
Man ya na.
I/NOM OR NO
I or no.

4.4.5. GT and VPE in Pro-form `Neither´

In the corpus, there were 55 relevant instances of `neither´ construction. GT translates

[15] The original sentence is negative, while the translation is apparently affirmative although it conveys the same negative meaning as the original sentence in this discursive context.

pro-forms with `neither´ as "نه (na) `no´ + subject", which is, in fact, an incomplete clause; the antecedent verb must fill the resultant gap (example (50)).

(50) *We fear nothing and neither will you.*

HT

ما از هیچ نمیترسیم و شما هم همینطور[16].

Ma az hich nemitarsim va shoma ham hamintor.

WE/NOM FROM/PRE NOTHING NEG+FEAR/PRS-1PL AND YOU/NOM-2PL TOO/PRF LIKEWISE/ADV

We fear-we from nothing and you too likewise.

We fear nothing and neither do you.

GT

ما هیچ ترس و نه شما.*

Ma hich tars van na shoma.

WE/NOM NOTHING FEAR/N AND NOR YOU/NOM-2PL

**We none fear and nor you.*

To sum up, English VPE in comparative pro-forms are translated into Persian pro-form structures by human translators. GT´s performance in dealing with these VPEs is summarized in table 5.

Table 5. GT of VPE after in pro-form structures

English Pro-forms	Cases	Persian Pro-forms (correct)	Persian Pro-forms (with errors)	Wrong Translation	Nonsense output
So	67	17.91%	26.87%	55.22%	–
Too	20	70%	–	30%	–
as well/also	23	–	100%	–	–
Either	10	–	–	–	100%
neither	55	–	–	–	100%
Total	175	14.86%	23.42%	24.57%	37.14%

4.5. Results

In English, all types of verbs are subject to VPE. VPE is very frequently constructed in the company of auxiliary verbs such as operators like various inflected forms of `be´, `do´, and `have´ or modal verbs (`can, may, will, would …´). Whereas, in Persian, it involves verbal groups in light verb constructions, verbs in simple present/past passive

[16] Refer to footnote 14.

voice, and verbs in company of modal verbs `can/could, may/might´ and `must´ (if it is translated as مجبور بودن (majboor boodan) [OBLIGED+BE/GR] `being obliged´).

The main problem of GT in dealing with English VPEs arises from translating the auxiliaries `do´, `be´, and `have´; consequently, GT produces noise in translation. GT also does not observe the agreement between subject and verb, in person and number. But since this issue is not directly related to VPE phenomenon, it is ignored in the analysis.

The next phase of the study was devoted to assess if the results of this study.

4.6. Assessment

The same corpus, which was used for the data analysis, was used in order to verify the representativeness of the data presented above against the corpus. The task was fulfilled on six random locations of the corpus (each containing 5% of the corpus, in total 183,607 sentences). In total, 1,094 cases of VPE instances were weighted up.

The results are summarized in table 6a and 6b (assessment of VPE after operators (`do, be, have´)) and table 7a and 7b (assessment of VPE after modals).

Table 6a. Assessment of VPE after operators (`do, be, have´)

Operators	Cases	Lexical verbs			Light verb constructions			
		VPE	Antecedent	Pro-verb	VPE	Antecedent	Pro-verb	NCE
Do	371	−	18.05%	13.21%	−	10.24%	13.21%	25.61%
Be	322	0.92%	15.52%	7.13%	−	7.44%	16.76%	20.55%
Have	43	−	39.53%	2.33%	−	9.30%	18.60%	23.26%
Total	736	0.40%	18.21%	9.92%	−	8.97%	15.08	23.23

Table 6b. Assessment of VPE after operators (`do, be, have´)

Lexical verbs/light verb constructions				
Operators	Confirming statements	Comparatives/Adverbials	Non-literals	Wrong/No Translation
Do	18.33%	1.35%	–	–
Be	31.06%	–	0.31%	0.31%
have	6.98%	–	–	–
Total	23.23%	0.68%	10.14%	0.14%

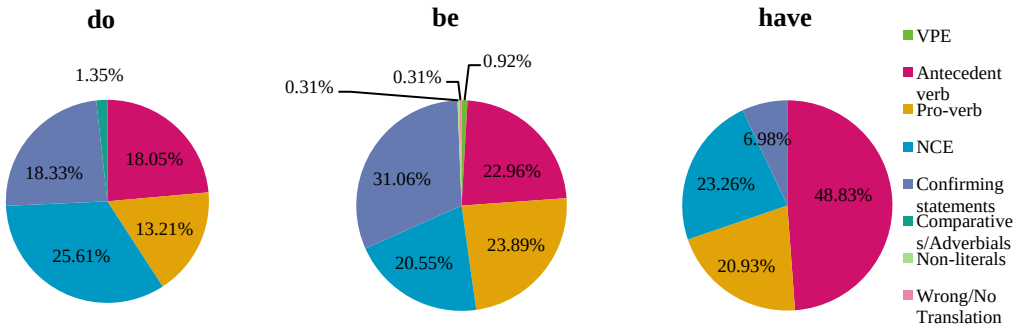

Figure 3. HT of VPE after operators (`do´, `be´, and `have´)

Comparing table 6 (a and b) and Fig 3 with table 2(a and b) and Fig. 2 proves that here again, the translators´ effort is more towards keeping the translation quite close to the original structure (75.81%) . VPEs after auxiliary `have´ were mostly recovered in Persian by the antecedent verb (48.83%). In the first analysis it seemed that the majority instances of VPE after auxiliary `be´ were in linguistic construction of a confirming answer to a previous statement; for which the adopted translation strategy was using the confirming statement ; however, here it is noticed this number is not as many as the previous analysis and like the previous analysis, the translation strategies, here, have been more or less evenly distributed among antecedent verb (22.96%), pro-verb (23.89%), NCE (20.55%), and confirming statements (31.06%). Again here, VPE after auxiliary `do´ were translated using quite similar number of occurrences of strategies (a little smaller than the first analysis): antecedent verb (18.05%), pro-verb (13.21%), NCE (25.61%), and confirming statement (18.33%). And finally, like the first analysis, a residual number of cases were translated using different constructions (comparative/adverbials) or were translated non-literally.

Table 7a. Assessment of VPE after modals

		Lexical verbs			Light verb constructions			
Modals	cases	VPE	Antecedent	Pro-verb	VPE	Antecedent	Pro-verb	NCE
can/may	141	31.21%	4.96%	10.71%	49.65%	2.13%	–	6.38%
will	119	–	23.53%	3.36%	–	19.33%	4.20%	23.53%
must/have to/should	98	9.18%	23.47%	5.10%	17.35%	18.37%	1.02%	25.51%
Total	358	14.81%	16.20%	2.79%	24.30%	12.29%	1.68%	17.32%

Table 7b. Assessment of VPE after modals

Lexical verbs/light verb constructions				
Modals	Confirming statements	Comparatives/Adverbials	Non-literals	Wrong/No translation
can/may	–	4.96%	–	–
will	23.53%	1.68%	0.84%	–
must/have to/should	–	–	–	–
Total	7.82%	2.51%	0.28%	–

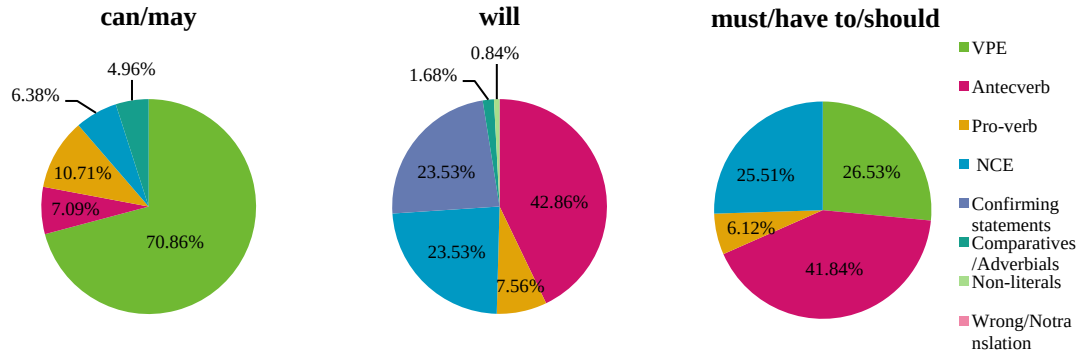

Figure 4. HT of VPE after modal verbs

Comparing table 7 (a and b) and Fig. 4 with table 3(a and b) and Fig. 2 indicate that, like the first analysis of data, it was proved that the translators have tried to keep the original structure in majority of cases (table 7a: 89.39%). Again, the structures with `can/may´ were mostly translated into Persian by keeping the VPE (70.86%), with considering the fact that here the frequency is much higher than the previous analysis (49.01%); the structures with `will´ were mostly translated using the antecedent verb (62.86%), not much different from the first analysis; and the structures with `most/have to/should´ kept the VPE for 26.52% of cases (with small difference from that of the first analysis), used the antecedent verb in translation for 41.84% of cases, and used NCE in 25.51%; that is, the results are quite similar with those of the previous analysis.

To sum up, the assessment data verify the representativeness of the obtained results of the study.

CHAPTER FIVE
CONCLUSION

This chapter presents the conclusions derived from the findings of this research based on the data from a subtitle corpus. Firstly, we outline the strategies adopted by human translators in dealing with English VPE. These strategies were then compared with GT in order to assess how it performs in translating those VPE instances. The systematization of these strategies may contribute to derive some regularity and general patterns which are expected to be useful information in improving the performance of English-Persian MT systems in future. Finally, some suggestions are presented to improve GT performance for this language pair which could motivate further research.

5.1. Overview

The present research adopted a descriptive corpus-based translation approach and focused on the patterns of translation of English Verb Phrase Ellipsis (VPE) into Persian. The goal was to find out the translation behavior.

An English-Persian, parallel corpus was used for this purpose. The corpus consists basically of informal conversations (dialogues) obtained from 1,600 movies´ subtitles, and it contains about 4 million words of each language. Some textual (mostly spelling) problems were found in the corpus, which hindered the detection of some instances of VPE. Thus, the English text was pre-processed and some faulty spellings were normalized.

Unitex finite-state tools were applied for defining certain generally formulated search patterns in order to detect the target English VPE. The extracted matches were then compared with their Persian counterparts.

5.2. Findings

Analysis of the Persian translations provided by the translators and the strategies utilized by them in dealing with English VPE is indicative of the fact that in those cases where Persian and English show some similar VPE constructions, the human translator keeps the translation quite close to the original text, even by retaining the ellipsis. However, in most cases, the elliptical forms are language-specific. In such cases, it is not possible to keep the ellipsis and the translator has to render the text in a non-elliptical form in order to provide the appropriate text, so to comply with Persian grammatical norms; that is, the gap resultant of VPE in English is usually recovered by the antecedent verb or replaced by a pro-verb.

When the two languages present similar construction, GT also produces a quite reasonable translation. However, in cases where Persian does not allow ellipsis, GT fails to recover the gap left by zeroed material in the source text. Auxiliary verbs also pose some specific problems, as GT translate them into light or lexical verbs.

Analysis of several examples and their translation by GT showed that this system proceeds in a sentence-by-sentence approach: given a sentence where VPE has occurred, with or without the preceding sentence where the antecedent is present, GT yields exactly the same result. This is probably one of the major drawbacks of this statistical MT system to deal with phenomena such as VPE, where some processing of anaphora is necessary.

Other observed problems in GT output are the lack of person and number agreement between the subject and verb, and also inconsistencies in the tense inflection of verbs. In Persian, being a highly inflected language, verbs are conjugated in person, number, and the tense, and the set of inflection forms for each verb is quite rich. Due to the differences among this language pair in this regard, GT produces noise. However, as these problems are not the subject of this study, they were ignored here and can be considered in future works.

In the next paragraphs, the main findings of this project are sketched, following the order by which the VPE phenomenon is presented in the body of this dissertation: First the ellipsis after auxiliary verbs is presented; then ellipsis after complementizer `to´ is discussed; and finally, the zeroing of the VP in the presence of pro-forms is referred to.

- *VPE with tense and modal auxiliaries:*

English VPEs occurring after auxiliaries `do´, `be´, and `have´ cannot be translated into Persian by keeping the ellipsis; therefore, the gap is usually filled by the antecedent verb or a pro-verb. However, if the English sentence carries a VPE after auxiliary `be´ and the sentence is translated into Persian using passive voice, then the VPE can be kept.

The Persian equivalents for many English verbs are light verb constructions. For these constructions, besides filling the gap with antecedent verb or pro-verb, another translation pattern is possible, namely, by keeping the light verb and omitting the nominal component.

GT produces distorted translations when dealing with English VPEs occurring after operators, since, as it was mentioned above, GT translate these auxiliaries literally and also because it does not recover the gap resulting from the English elliptical sentence.

Persian allows keeping the gap produced by the English VPE when this involves the modal verbs `can´, `may´, or `must/have to´, if they are translated as مجبور بودن (majboor boodan) [OBLIGED+BE/GR]). Naturally, the recovering of the gap by the antecedent verb or by a pr-verb is always possible in this case, as well as the use of a light verb with the zeroing of the nominal component. However, after the modal verbs `will´, `should´, and `must/have to´, if they are translated as باید (bayad), the English VPE cannot be kept in Persian. The gap must then be recovered. In the case of light verb constructions the light verb is kept but the nominal component is zeroed.

For VPEs after modal verbs, GT performs quite acceptably but only after modal `can´;

however it fails in dealing with other modal verbs.

- *VPE after complementizer `to´:*

English VPEs occurring after infinitival complementizer `to´ (654 cases) are mostly translated by filling the gaps with the antecedent verb (513 cases); in 127 cases, the ellipsis was kept; and in a residual number of cases (12 instances) the ellipsis was filled by a pro-verb.

GT retains this kind of ellipsis in all cases. Since it does not fill the gap resulting from English VPE, the output is often unnatural even if it may be interpretable in most cases.

It is worth mentioning that in the English corpus, the verbs involved in the VPE and preceding the infinitival complementizer `to´ are not very numerous. In this particular situation, we found, from the most to the least frequent one, *`want, be supposed, used, try, like, mean, tend, expect, pretend, need, seem, wish, dare, commit, attempt´*. This observation can be taken as an advantage in defining some rules for improving MT performance regarding VPE with these verbs when followed by complementizer `to´.

- *VPE with pro-form constructions:*

English VPEs with pro-form structures with `so/too/as well/neither/either´ (175 cases) were, for the most part, translated into Persian using also pro-forms; thus, keeping the ellipsis. Only in a residual number of cases the gap was filled with the antecedent verb or replaced by a pro-verb.

From among 67 cases of VPE in pro-form structure with `so´, only 12 cases were properly translated by GT using an adequate pro-form structure. The system, however, failed in dealing with the other cases either by translating the structure into an adverbial clause in which `so´ played the role of a conjunctive adverb بنابراین (banabrin) signifying `therefore´; or by translating the auxiliary verbs and producing an unacceptable output.

From the collected evidence, it was not possible to discover why GT only performs properly in some cases.

23 instances of VPE using pro-forms `as well´ and `also´ were translated into Persian equivalent pro-forms but involving the incorrect translation of auxiliary verbs.

From 20 VPE instances in pro-forms with `too´, 14 cases were translated by using pro-forms; the rest (6 cases) `too´ were translated to the degree adverb بیش از حد (bish az had) signifying `excessively´.

GT, in treating with all instances of VPE in pro-form structures with `either´ and `neither´, produced nonsense outputs.

To sum up, it seems that GT, in dealing with this particular type of VPE, mostly produces inadequate translations.

Besides the above-mentioned patterns of translation, one construction was observed that could not follow the same translation strategies: VPE in adverbial and comparative clauses; for this case the structure is reduced to Persian adverbial and comparative phrases.

It should be noted that in some few cases the translators do not feel obliged to keep the same structure as the original; while keeping the content, they did not keep the original wording. This tendency may be justified for the purpose of making the text more natural and adequate, especially in conversational (oral) situation of subtitles; and for creating diversity and expressiveness; or even for conveying the ironic sense of the sentence. This approach, however, is only adopted in a few numbers of instances in the corpus. Therefore, it can be claimed that the patterns here studied were, for the most part, quite close to those of the original structures.

5.3. Final Remarks

The results indicate that Persian human translator dealing with English VPE predominantly adopts the strategy of recovering the zeroed verb from its previous occurrence in discourse. Naturally, in some cases, instead of a verb, a pro-verb is used. For light verb constructions in Persian, the light verb is retained and the nominal component is zeroed. This ellipsis is, in fact, a sub-category of VPE.

This general behavior, however, depends on the auxiliary verb used in the source language. Differences in the auxiliary verb used in English VPE have a relevant bearing on the choice of the strategies the human translator adopted. For instance, the translation strategy for VPE after modal verbs `can´, `may´, and `must/have to´, and for VPE in pro-form structures is mostly keeping the ellipsis in the text, while for VPE after other modal verbs and after operators, the strategy is filling the gap by the antecedent verb or a pro-verb.

As a statistical-based MT system, GT does not take into consideration the discourse previous to the sentence under processing. The system, therefore, is incapable to recover the gap induced by English VPE, this results in incorrect translation output as it has also been confirmed by the analysis.

The comparison between HT and GT of Persian texts indicates that a stronger effort should be invested in an anaphora resolution module, particularly, for certain English VPE patterns: those involving auxiliary verbs` do´, `be´, `have´, and `will´, and those after complementizer `to´.

The use of comparative pro-forms also yields much noise in GT, probably because the English VPE has not been adequately parsed or identified. While the VPE can, in most cases, be kept in the Persian translation, the adequate choice of the pro-form depends on the precise identification of a VPE in the first place.

The findings of this study may help devise better performing strategies for English-Persian MT systems, namely, by highlighting the relevance of an anaphora resolution module prior to the MT of this language pair.

5.4. Suggestions for Further Research

The scope of this study was limited to those patterns that would be most easily captured in the available corpus. An evident development of this work should use richer linguistic information to help spotting those instances of VPE that were not considered here.

In this study, the situational context of text was conversational by nature; conversational texts have a structure of their own quite different from other texts. In conversational texts, there are so many linguistic phenomena, such as shift turning, unconcluded sentences, texts appearing in gestures rather than words, etc. that can affect the VPE phenomenon. It is not obvious that findings on colloquial (oral) discourse can be straightforwardly transposed to other discursive situations and genres. Extension of this study to other textual types may illuminate better the general issues pertaining to English-Persian VPE translation. We expect that increasingly available, new parallel corpora will make this extension possible in the near future.

References

Abedi, M. (2004) *Small Corpora in Translation Studies: Comparing Thematic Organization in Two Languages English and Persian* (M.A. Thesis), Allameh Tabatabai University, Tehran.

Anvari, H, and Givi, H. (1999) *Dastoore Zabane Farsi 2* (Persian Grammar). Tehran: Fatemi Press.

Asher, N., Hardt, D. and Busquets, J (2001) Discourse Parallelism, Ellipsis and Ambiguity. *Journal of Semantics*, **18**(1), pp. 1-25.

Aston, G. (1999) Corpus Use and Learning to Translate. *Textus*, 12, pp. 289-314.

Baker, M. (1993) Corpus Linguistics and Translation Studies: Implications and Applications. *In* Baker, M. Francis, G., and Tognini-Bonelli, E. (eds). *Text and Technology: in Honor of John Sinclair*, Amsterdam/Philadelphia, John Benjamins, pp. 233-250.

Baker, M. (1996) Corpora in Translation Studies: An Overview and some Suggestions for Future Research, *Target*, **7**(2), pp. 223-243.

Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics - Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Bos, J. (2005) Towards Wide-coverage Semantic Interpretation. In *Proceedings of IWCS-6*, pp-42-53, Tilburg, The Netherlands.

Bowker, L. (2000) A Corpus-based Approach to Evaluating Student Translations. *The Translator: Evaluation and Translation* (Special Issue), **6**(2), pp. 183-210.

Bowker, L. (2001) Towards a Methodology for a Corpus-based Approach to Translation Evaluation. *Meta*, **46**(2), pp. 345-364.

Busquets, J. (1997) A Discourse-Based Approach to VPE: A Survey. *Rapport Interne IRIT/97-48-R, IRIT*. Universite Paul Sabatier, Toulouse.

Church, K. and Gale, W. (1991) Concordances for Parallel Text. In *Using Corpora*: *Proceeding of the 7th Annual Conference for the New OED and Text Research*, Oxford, pp. 40-62.

Clark, S. and Curran, J. (2004) Parsing the WSJ using CCG and Log-linear Models. In *Proceedings of 42nd Annual Meeting of Association for Computational linguistics (ACL ´4),* pp. 104-111, Barcelona, Spain.

De Vries, D. (2009*) A Semantic Approach to Antecedent Selection in VP Ellipsis* (M.A. Thesis), University of Groningen.

Farahzad, F. (2000) *Linguistics Applied to Translation Studies*. Tehran: Allame Tabatabai´s University Press.

Farshidvar, K. (1997) *Jomle va Tahavole an dar Zabane Farsi* (Sentence and its Evolution in Persian Language). Tehran: Fatemi Press.

Garner, B. (2004) *A dictionary of Modern Legal Usage* (2nd ed.). Oxford University Press Inc.

Ge, N., Hale, J. and Charniak, E. (1998) A Statistical Approach to Anaphora Resolution. In *Proceedings of the 6th Workshop on Very Large Corpora*, Harcourt Brace, pp. 161-171.

Granger, S., Lerot, J. and Petch-Tyson, S. (2003) *Corpus-based Approaches to*

Contrastive Linguistics and Translation Studies. Amsterdam/Atlanta: Rodopi.

Guangrong, D. (2009) The Applications of Keywords and Collocations to Translation-Studies and Teaching---A Tentative Research on the Parallel Corpus of the 17th NCCPC Report. *Translation Journal,* **13**(3) [online: no pagination]. [Accessed 28 Feb 2010]. Available at: <http://accurapid.com/journal/49corpus.htm>.

Haase, F. (2008) *States of the Common and the Unique: An Introduction to a General Functional Communication Theory*. [online]. [Accessed 5 April 2010]. Available at: <http://ssrn.com/abstract=1100062>.

Halliday, M. and Hasan, R. (1976) *Cohesion in English*. London: Longman.

Hardt, D. (1997) An Empirical Approach to VP Ellipsis. *Computational Linguistics*, **23**(4), pp. 525-541.

Harris, Z. (1991) *A Theory of Language and Information*: A Mathematical Approach. Oxford: Clarendon Press.

Harris Z. (1955) *Co-occurrence and Transformation in Linguistic Structure*. *Language*, **33**(3), pp. (283-340), reprinted in Harris, Z. (1981). Papers on Syntax (ed. Henry Hiz). Dordrecht/Boston: D. Reidel Pub.Co.

Hashemianjazi, H. (1989) *A Contrastive Analysis of English and Persian Ellipsis* (M.A. Thesis), University of Shiraz, Shiraz (Iran).

Holmes, J. S. (2000). The name and Nature of Translation Studies. In L. Venuti (Ed.), *The Translation Studies Reader*. London, United Kingdom: Routledge, pp. 172-185.

Huddlestone, R. and Pullum, G. (2002) *The Cambridge Grammar of the English Language*. Cambridge, Uk: Cambridge University Press.

Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Kamp, H. and Reyle, U. (1993) *From Discourse to Logic: An Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and DRT.* Kluwer: Dordrecht.

Karimi- Doostan G. (1997) *Light Verb Constructions in Persian* (PhD. Thesis), University of Essex, England.

Kerdar, A. (2005) *A Corpus-Based Study on Translation of the English Passive Voice into Modern Persian Literary Texts* (M.A. Thesis), University of Allameh Tabatabai, Tehran (Iran).

Khanlari, T. (2001) *Dastoor Zaban Farsi* (Persian Grammar). Tehran: Tus Press.

Lambert, J. and Van Gorp, H. (1985). On Describing Translations. *In* Hermans, T. (ed.), *The Manipulation of Literature. Studies in Literary Translation*. London/Sydney: Croom Helm, pp. 42-53.

Lappin, S. (1996). The Interpretation of Ellipsis. *In* Lappin, S. (ed.), *Handbook of Contemporary Semantic Theory*. Oxford: Blackwell. pp. 145-175.

Lappin, S. (2005) A Sequenced Model of Anaphora and Ellipsis Resolution. *In* Branco, A. McEnery, A., and Mitkov, R. (eds.*). Anaphora Processing: Linguistic, Cognitive, and Computational Modeling*. Amsterdam: John Benjamins, pp. 3-16.

Lappin, S. and Benmamoun, E. (eds.). (1999) *Fragments: Studies in Ellipsis and Gapping*. New York: Oxford University Press.

Lappin, S. and Leass H. (1994) An Algorithm for Pronominal Anaphora Resolution. *Computational Linguistics*, **20**(4), pp. 535- 561.

Lappin, S. and McCord, M. (1990) Anaphorical Resolution in Slot Grammar. *Computational Linguistics*, **16**(4), pp. 197-212.

Leech, G. (1992) 100 Million Words of English: The British Natural Corpus. *Language Research*, **28**(1), pp. 1-13.

Lobeck, A. (1995). *Ellipsis*. Cambridge: Cambridge University Press.

Maduro, R. and Carvalho A. (2002) Syntactic Analysis for Ellipsis Handling in Coordinated Clauses. In *Proceedings of the 16th Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence.* Springer-Verlag, London, pp. 397-406.

Malmkjær, K. (1998) Love thy Neighbor: Will Parallel Corpora Endear Linguists to Translators? *Meta*, **43**(4), pp. 534-541.

McCarthy, M. and Carter, R. (2001) Ten Criteria for a Spoken Grammar. *In* Hinkel, E. and Fotos, S. (eds.). *New Perspectives on Grammar Teaching in Second Language Classrooms*. Mahwah, NJ: Lawrence Erlbaum, pp. 51-75.

McCord, M. (1990) Slot Grammar: A System for Simpler Construction of Practical Natural Language Grammars. In *Proceedings of the International Symposium on Natural Language and Logic*. London: Springer Verlag, pp.118-145.

McEnery, A. and Xiao, Z. (2005a) Help or Help to: What do Corpora Have to Say? *English Studies*, **86**(2), pp. 161-187.

McEnery, A. and Xiao, Z. (2005b) A corpus-based Approach to Tense and Aspect in English-Chinese Translation. In J. Schwitalla and W. Wegstein (Hg.) *Korpuslinguistik deutsch: synchron - diachron - kontrastiv* . Niemeyer: Tübingen, pp. 27-50.

McEnery, A., Xiao, Z. and Tono, Y. (2006) *Corpus-Based Language Studies - an*

Advanced Resource Book. London: Routledge.

McEnery, T. and Wilson, A. (2001) *Corpus Linguistics: An Introduction*. (2nd ed.) Edinburgh: Edinburgh University Press.

Marcus, M., Kim, G., Marcinkiewicz, M., MacIntyre, R., Bies, M., Ferguson, M., Katz, K. And Schaberger, B. (1994a) The Penn Treebank: Annotating Predicate Argument Structure. In *Proceedings of the Human Language Technology Workshop*. Morgan Kaufmann, San Franscisco.

Marcus, M., Santarini, B. and Marcinkiewicz, M. (1994b) Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, **19**(2), pp. 313-330.

Marinai, E., Peters, C., and Pichhi, E. (1991) Bilingual Reference Corpora: A System for Parallel Text Retrieval. In *Using Corpora*: *Proceedings of the 7th Annual Conference of the UW Centre for the New OED and Text Research*, Oxford, pp. 63-70.

Marinai, E., Peters, C., and Pichhi, EE. (1992) Bilingual Reference Corpora: Creation, Querying, Applications. *In* Kiefer, F. Kiss, G, and Pajzs, J. (eds), *Papers in Computational Lexicography: Complex 92*, Budapest, Linguistics Institute, Hungarian Academy of Sciences, pp. 221-228.

Martin, A. E., & McElree, B. (2008) A Content-Addressable Pointer Mechanism Underlies Comprehension of Verb Phrase Ellipsis. *Journal of Memory and Language,* **58**, pp. 879-906.

Megerdoomian, K. (2000) Persian Computational Morphology: A Unification-based Approach. *Memoranda in Computer and Cognitive Science Report MCCS- 00- 320*, Computing Research Lab, New Mexico State University. [online]. [Accessed 16 Nov 2010]. Available at: <http://citeseer.ist.psu.eda/megerdoomian00Persian.html>.

Mitkov, R. (1995) An Uncertainty Reasoning Approach for Anaphora Resolution. In *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLPRS95)*, pp. 149-154.

Mitkov, R. (1999) Multilingual anaphora Resolution. *Machine Translation*, **14**(3-4), pp. 281-299.

Mitkov, R. (2002) *Anaphora Resolution*. London: Longman.

Mitkov, R., Boguraev, B., and Lappin, S. (eds.) (2001) Introduction to Special Issue on Computational Anaphora Resolution. *Computational Linguistics*, **27**(4), pp. 473–477.

Mitkov, R., Evans, R. and Orasan, C. (2002) A New, Fully Automatic Version of Mitkov's Knowledge-poor Pronoun Resolution Method. In *Proceedings of* CICLing-*2000*, Mexico City, Mexico, pp. 168-173.

Mohammadian, G. (2007) *A Corpus-based Study on Norm-Governed Strategies of Dialogue Translation in Persian Translated Novels* (M.A. Thesis), Allameh Tabatabai University, Tehran (Iran).

Mohebbi, B. (1998) *A Contastive Analysis of Ellipsis in English and their Translation Equivalents in Persian* (M.A. Thesis), University of Shiraz, Shiraz (Iran).

Mosavie-Miangah, T. (2009) Constructing a Large-Scale English-Persian Parallel Corpus**.** *Meta*, **54**(1), p. 181-188.

Munday, J. (2001) *Introducing Translation Studies: Theories and Applications*. London: Routledge.

Murguia, E. (2004) *Syntactic Identity and Locality Restrictions on Verbal Ellipsis.* (PhD. Thesis) University of Maryland.

Nielsen, L. A. (2005) *A Corpus-based Study of Verb Phrase Ellipsis* (PhD. Thesis), University of London.

Normohammadi, K. (1988) *A Contrastive Analysis of Cohesion in English and Persian* (M.A. Thesis), Shiraz University, Shiraz, (Iran).

Palomar, M., Moreno, L., Peral, J., Muñoz, R., Ferrández, A., Martínez-Barco, P. and Saiz-Noeda, M. (2001) An Algorithm for Anaphora Resolution in Spanish Texts. *Computational linguistics,* **27**(4), pp. 545-567.

Paumier, S. (2003) *De la Reconnaissance de Formes Linguistiques à I'analyse Syntaxique* (PhD. Thesis), Univ. MLV, Marne-la-Vallée, Paris.

Paumier, S. (2008) *Unitex 2.0 User Manual*. Paris: Université de Marne-la-Vallée (UMLV).

Peters, C. and Picchi, E. (1998) Bilingual Reference Corpora for Translators and Translation Studies. *In* Bowker et al. (eds.), *Unity in Diversity? Current Trends in Translation Studies* Manchester: St. Jerome Publishing, pp. 91-100.

Pilevar, M. (2010) *Tehran English Persian Parallel Corpus*. [online]. [Accessed 4 April 2010]. Available at: <http://ece.ut.ac.ir/nlp/resources.htm>.

Pilevar, M.T. and Feili, H. (2010) Persian SMT: A First Attempt to English-Persian Statistical Machine Translation, in *Proceedings of 15th International Conference of Statistical Analysis of Textual Data (JADT),* Rome, pp. 1101-1112.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Ross, J. R. (1967) *Constraints on Variables in Syntax* (PhD. Thesis), Cambridge,

Massachusetts: Massachusetts Institute of Technology (MIT).

Shahrestanee, Z. (2007) *A Corpus-based Study of Units of Translation in English-Persian Literary Translations* (M.A. Thesis), Mofid University, Qum (Iran).

Schäffner, C. (2000) Parallel Texts in Translation Studies. In Bowker et al. *Unity in Diversity: Current Trends in Translation Studies*. Manchester: St. Jerome Publishing, pp. 83-90.

Schwarz, B. (2004) Indefinites in Verb Phrase Ellipsis. *Linguistic Inquiry*, **35** (2), pp. 344-353.

Shariati, M. J. (2001) *Dastoor Zaban Farsi* (Persian Grammar). Tehran: Asatir Press.

Shlesinger, M. (1998) Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies. *Meta*, **43**(4): 486-493.

Sinclair, J. (1991) *Corpus, Concordance, and Collocation*. Oxford: Oxford University Press.

Teubert, W. (2003) Writing, Hermeneutics, and Corpus Linguistics. *Logos & Language*, **4**, pp. 1-17.

Tiedemann, J. (2007) Improved Sentence Alignment for Movie Subtitles, in *Proceedings of 12th Recent Advances in Natural Language Processing (RANLP 2007)*. Borovets, Bulgaria, pp. 585-588.

Tohidian, I. (2008) Corpus-based Translation Method. *California Linguistic Notes*, **33**(1) [online]. [Accessed 11 April 2010]. Available at: <http://hss.fullerton.edu/linguistics/cln/W08%20pdf/tohidian-Corpus-based.pdf>.

Toosarvandani, M. (2009) V-Stranding VPE: Ellipsis in Farsi Complex Predicates.

Blackwell Publishing Ltd, *Syntax*, **12**(1), pp. 60-92.

Toury, G. (1994) *Descriptive Translation Studies-and Beyond*, Amsterdam: John Benjamins.

Varantola, K. (2003). Translators and Disposable Corpora. In Zanettin, F. Bernardini, S. and Stewart, D. (eds.) *Corpora in Translator Education*, Manchester: St Jerome, pp. 55-70.

Williams, J, and Chesterman, A. (2002) The *Map: A Beginner's Guide to Doing Research in Translation Studies*. Manchester: St. Jerome Publishing.

Yamamoto, K. and Sumita, E. (1998) Feasibility Study for Ellipsis Resolution in Dialogues by Machine-Learning Technique. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL)*. Stroudsburg: Association for Computational Linguistics, **2**, pp. 1428-1435.

Yarmohammadi, L. (1995) *Fifteen Articles in Contrastive Linguistics and the Structure of Persian: Grammar, Text, and Discourse*. Tehran: Rahnaman.

Yan, D., Noël, D. and Wolf H. (2010) Patterns in Metaphor Translation: a Corpus-based Case Study of the Translation of FEAR Metaphors between English and Chinese. In Xiao, R. (2010) (ed.). *Using Corpora in Contrastive and Translation Studies*. Newcastle: Cambridge Scholars Publishing, pp. 40-61.

Zhang, M. (2002) Using Corpus for Investigating the Style of a Literary Translator-Introducing and Commenting on Baker's New Research Method [J]. *Journal of PLA Foreign Languages University*, **25**(3), pp. 54-57.