Are Translation Universals Really Universal? A Corpus-based Study of Translational Expository Persian

Morteza Taghavi
Ferdowsi University of Mashhad, IRAN
mortezataghavi@mail.um.ac.ir

Mohammad Reza Hashemi
Ferdowsi University of Mashhad, IRAN
hashemi@um.ac.ir

ABSTRACT

Central to descriptive Translation Studies is the concept of Translation Universals (TUs), referring to linguistic features idiosyncratic to translation, typically not occurring in original texts and not influenced by a given language pair. Over the past three decades, almost all literature about TUs is on Western languages, especially English as translated from/to other related European languages, and other languages have been ignored. If features of translational language are to be generalized as TUs, supporting evidence from non-European languages should be provided. Using Chesterman's (2004) categorization of S-universals and T-universals, the present corpus-based study sets out to investigate T-universals of simplification and explicitation in a comparable corpus of translational and original Persian expository texts. Aiming at finding distinctive lexical and syntactic features of translational Persian, this study raises intriguing questions regarding the presence of universal features in translations as none of the results regarding the features addressed were in line with previously proposed T-universals.

KEYWORDS: translation universal; comparable corpus; translational Persian; T-universals; simplification; explicitation

1. Introduction

One of the most significant topics within descriptive Translation Studies is Translation Universals (TUs), first clearly articulated by Mona Baker (1993). The term refers to typical features of translational language that differentiate it from other linguistic variants. As pointed out by Silvia Hansen and Elke Teich (2001), "it is commonly assumed in translation studies that translations are specific kinds of texts that are different not only from their original source language (SL) texts,

but also from comparable original texts in the same language as the target language (TL)" (44). In order to examine such claims, corpora have been a popular tool since the convergence of corpus-based empirical methodology and linguistic studies, including Translation Studies, during the 1990s. As Sara Laviosa (1998a) notes, "the corpus-based approach is evolving, through theoretical elaboration and empirical realization, into a coherent, composite and rich paradigm that addresses a variety of issues pertaining to theory, description, and the practice of translation" (474). Over the past three decades, a significant number of studies – almost all on Western languages, especially English – have addressed TUs (see *e.g.* Blum-Kulka 1986; Øverås 1998; Olohan and Baker 2000; Teich 2001), all of which more or less provide evidence of the linguistic features that are peculiar to translational language.

Andrew Chesterman (2004) distinguishes between S-universals and T-universals. The former refers to "universal differences between translations and their source texts" (Chesterman 2004:39), while the latter covers the universal differences between translations (target texts) as compared to non-translated, native TL texts. In order to investigate them with a corpus-based methodology, S-universals require a parallel corpus of source and target texts and T-universals require a comparable corpus of translated and native target texts. In particular, potential S-universals include: 'translations tend to be longer than their source texts', 'explicitation', 'sanitization' (reduced connotative meaning), 'later translations tend to be closer to the source text', etc. On the other hand, potential T-universals may comprise: 'simplification', 'conventionalization', 'untypical lexical patterning', 'under-representation of TL-specific items', etc. (Chesterman 2004:40). Although Chesterman categorizes universals into two groups, some features may overlap.

Almost all literature on TUs has investigated Western languages; English outbalances all other languages in terms of having been studied as the most recurrent language translated to or from other European languages. Research on other languages is scarce. Only a very small number of works on non-Western or non-European languages (e.g., Xiao 2010; Xiao and Hu 2015, on the Chinese language) have addressed the topic. Moreover, research on S-universals outweighs work on T-universals. Hence, such claims as the existence of 'universal' features, in the strict sense of the word 'universal', is highly debatable, unless they are scrutinized in other languages, especially

those that are different from English in terms of word order, syntactic structures, stylistic features and the like.

Investigation of TUs has also been neglected to a great extent in studies on the Persian language. Additionally, all we know about TUs in translational Persian, however little, is based on small-scale case studies limited to literary translations as compared to their source texts (S-universals) (e.g., Beikian et al. 2013; Vahedi Kia and Ouliaeinia 2016; see section 2.3). To fill this gap, this study focuses on examining T-universal hypotheses through comparing English-to-Persian translations of non-literary expository texts with comparable original texts to identify distinctive features of translational language. The results of the investigation will cast light on what happens during the translation process that makes translational language idiosyncratic, or the 'third code', to quote William Frawley (1984). In particular, the present study attempts to investigate the T-universals of simplification and explicitation by exploring a number of prominent linguistic features of translational Persian in a comparable corpus of translated and native Persian texts drawing on non-literary texts. To this end and as an instance of non-literary writings, a medium-sized corpus consisting of two sets of expository academic and general texts belonging to Humanities fields, namely psychology and sociology, were employed to examine whether the previous propositions on two of the most prominent TUs, simplification and explicitation, are applicable to a non-Western language.

2. Background and Theoretical Framework

2.1 The Search for Translation Universals Around the World

The search for Tus dates back to the mid-nineties, where this topic led to a surge of interest among researchers, especially since the emergence of corpora as a research tool in Translation Studies. One of the first studies on Tus was Frawley's (1984) discussions on translations as having a 'third code' which is distinct from both the source and target language codes. Although the third code refers to the concept of *tertium comparationis*, i.e., the similar quality shared by the two elements in a comparison, here, the emphasis is on the idiosyncrasy of translational language. It was not until Baker's (1993) seminal paper that "the idea of linguistic translation universals found a place at the center of discussion in translation studies" (Mauranen and Kujamäki 2004:1). Translation

universals refer to the specific linguistic characteristics of translations that exist only by virtue of their being translated and which are the result of the translation process.

A few attempts have been made to orchestrate the variety of TU hypotheses, the most well-known of which is Chesterman's (2004). In a chapter entitled 'Beyond the Particular', he examines three main ways of moving 'beyond the particular' (making generalizations about translation), namely "traditional prescriptive statements, traditional critical statements, and the contemporary search for universals in corpus studies" (Chesterman 2004:33). In search of Tus within the third route away from the particular, Chesterman (ibid.) divides universals into two general categories of S-universals and T-universals. Although some universals would seem to fall naturally under the S-universals category, there are some – like explicitation (Xiao 2010:9) or simplification – that have been studied within the category of T-universals, depending on the research focus and data under scrutiny. For example, we can either examine whether translators simplify language in translated texts compared to their sources, or compare translations with comparable native texts to see possible manifestations of simplification.

Taking a closer look at the existing literature on Tus shows that the research carried out on S-universals outweigh the studies on T-universals. In the past three decades, a considerable amount of literature has been published comparing translations with their corresponding source texts (e.g., Blum-Kulka 1986; Baker 1993; Toury 1995; Øverås 1998; Olohan and Baker 2000; Teich 2001, 2003; Mauranen 2007; Xia and Hu 2015; Molés-Cases 2019). On the other hand, T-universals have also been explored. It is noteworthy that the investigation of universals has not been limited to translations; a number of researchers have also investigated universals of interpretations (for example, Gumul 2006; Kajzer-Wietrzny 2012; Morselli 2018). In addition, a new line of research has started to probe into cognitive and psycholinguistic features of translation universals (Zasiekin 2019). The focus of this study is on comparing translational language with comparable native texts (i.e., T-universals), excluding studies on S-universals.

2.2 The Search for T-Universals

In one of the first studies on T-universals, Laviosa-Braithwaite (1996) investigated simplification via a comparable corpus of translated and original English texts in two genres: newspaper articles

and narrative prose. She also employed type-token ratio (TTR), lexical density and mean sentence length as areas for possible manifestation of simplification. Her study revealed four 'core patterns of lexical use' which can be evidence of lexical simplification in translated versus original texts, namely 1) relatively lower proportion of lexical words versus grammatical words, 2) relatively higher proportion of high frequency versus low frequency words, 3) relatively greater repetition of the most frequent words and 4) less variety in the words most frequently used. Václav Cvrček and Lucie Chlumská (2015) performed a study on lower lexical richness, as a manifestation of simplification, in a comparable corpus of translated and non-translated Czech fiction and academic texts (from fields such as law, medicine, history, music, chemistry, etc.). However, they neither resorted to the widely used TTR, because it is sensitive to text size, nor the improved version, i.e., standardized type-token ratio (STTR), because it takes no notice of intratextual variability and dynamics. Instead, they decided to use another version called zTTR which works based on the comparison of the actual TTR with referential values, taken from a large reference corpus, to reflect both the size of a text and its text type. Nevertheless, the study yielded a similar conclusion to Laviosa-Braithwaite (1996) as translated Czech texts had a tendency to show a slightly less diverse lexicon compared to non-translated Czech texts.

However, the simpler language of translated text as compared to that of an original, non-translated text, purportedly, is not limited to the word-level; it may be manifested at syntactic and/or stylistic levels as well. For instance, Laviosa (1998b) inquired into a comparable corpus of English narrative prose and reported that mean sentence length in translated language is significantly greater than in non-translated language. Similarly, Richard Xiao and Ming Yue (2009) investigated Chinese, which is one of the most investigated non-English languages with regard to Tus, especially T-universals (e.g., Chen 2006; Xiao and Hu 2015). Their study is in line with Laviosa (1998b) as they observed that translated Chinese fiction shows a significantly greater mean sentence length than native Chinese fiction.

Contrary to these findings, there is a disagreement on the validity of mean sentence length as a sign of simplification. For example, Gloria Corpas Pastor et al. (2008) adopted an NLP approach to test simplification and convergence in comparable corpora of translated and non-translated Spanish. With the help of language processing tools, they analyzed the corpora in relation to a

variety of lexical, grammatical, and stylistic characteristics. The results supported the hypothesis of simplification of translated texts by displaying their lower lexical density, albeit "this [was] not true with regard to the sentence length and the use of simple vs. complex sentences, and texts produced by non-professional translators [did] not seem to possess such simplification traits" (Corpas Pastor et al 2008:7).

There is also a bulk of research supporting the validity of mean sentence length. In another study, Iustina Ilisei et al. (2009) demonstrated that lexical richness and mean sentence length are the most outstanding indicative features of the simplification hypothesis. They developed a supervised learning system to distinguish between translated and non-translated texts with high accuracy. The study benefited from three Spanish comparable corpora – two comparable corpora made of medical texts and one technical text – and extracted 21 language-independent features to be utilized by their learning system in distinguishing between translated and non-translated texts. By analyzing the various classifiers that could be indicators of simplification, they concluded that lexical richness and mean sentence length, among others, are the most salient features that characterize translated vs. original texts.

A number of other studies have further demonstrated evidence for explicitation. This hypothesis was first formulated by Shoshana Blum-Kulka (1986) who argued that translations tend to exhibit more cohesive markers than non-translations. As an example, a study by Maeve Olohan and Mona Baker (2000) on translational English in TEC (Translational English Corpus) and native English in the BNC (British National Corpus) indicated that the *that*-connective is far more frequent in TEC than BNC. However, thus far, this manifestation of explicitation is not limited to the higher frequency of connectives. Olohan (2004) proposed that examining how a text uses moderating words (like *quite* and *rather*) can be a way of investigating explicitation. She examined intensifiers such as *pretty, rather, quite* and *fairly* and noted that these four items are markedly less frequent in translated English fiction than in native English fiction. This observation leads her to connect moderation to explicitation by explaining that "translators may remove or downplay elements of 'moderation', perhaps as part of a (non-deliberate) process of disambiguation or explicitation" (Olohan 2004:142).

2.3 The Search for Universals in Translational Persian

What we know about the possible presence of Tus in translational Persian is mostly based on studies limited in one way or another. For example, Ali Beikian et al. (2013) performed a corpus-based study on explicitation as a possible S-universal and concluded that translated texts used more explicit connectives than their corresponding source texts. However, their data were limited to just one-third of an English book with its translation and the focus of the study was on comparing a source text with its target text, which many other studies had done before. Mehdi Vahedi Kia and Helen Ouliaeinia (2016) show the relative presence of explicitation in English translations of modern Persian literary works, but with only 80,000 source-text words. In another example, Abbas Ali Ahangar and Seyyedeh Nazanin Rahnemoon (2019) examined the level of explicitation of reference in the published and Google translations of medical texts. Again, only one English book, its two Persian translations and the machine translation version were used as the data. Another relevant issue is that of 'methodology', with many studies using manual investigation and not benefiting from corpus investigation tools. These and many other studies on Persian share the same areas of focus, leading the literature to leave other areas untouched. Such explored and investigated areas in the literature on Persian can be categorized as follows:

- direction being restricted to comparison of source with target text(s) (addressing only S-universals)
- lack of variety in source language (almost all studies feature English as the source)
- universals (all studies are on the four recurrent features of translation proposed by Baker (1996))
- genre (almost all studies focus on literary texts)
- size (very small-scale studies, mainly on selected parts of one or two books)
- methodology (using manual investigation and not benefiting from corpus investigation tools)
- source of data collection (all data from books and published works, ignoring online translated materials available)

Surprisingly, except for two studies (Esteki et al. 2010; Alibabaee and Salehi 2012), we found no study examining T-universals using a comparable corpus of translated and native Persian texts. In

their study which aimed at testing the simplification hypothesis, Azadeh Esteki et al. (2010) developed a corpus of Persian translated and untranslated economic texts. They selected three features indicating simplification and the results showed that translated Persian texts were only simpler in terms of low sentence length and not in terms of high type-token ratio and low lexical density. In another study on political texts, Ahmad Alibabaee and Zahra Salehi (2012) found that lexical density and type/token ratio in political translated texts were higher than those of the political non-translated texts, but political translated texts reported lower mean sentence length. Additionally, regarding both T- and S-universal research, there are only three studies with data other than literary texts, namely Esteki et al. (2010) on economic texts, Alibabaee and Salehi (2012) on political texts, and Ahangar and Rahnemoon (2019) on medical texts. As mentioned in Section 2.1, new lines of research on Tus have emerged. However, it can be seen that little is known about Persian and the present study seeks to address some of these persistent research gaps.

2.4 Arguments for and Against Universality of Universals

Over the past two decades, scholars have argued for and against universals of translations from two different viewpoints. Some scholars are against the very idea of making general claims about translations (e.g., Tymoczko 1998; Malmkjær 2007; House 2008), whereas Gideon Toury (2004) believes that they are valuable for their explanatory power and, as Chesterman (2004) puts it, another way to make generalizations about shared features of translations.

Regarding TUs, it is important to make a distinction between 'universals' or laws, and norms. As Sari Eskola (2004) points out, norms are mainly prescriptive and culture-bound, while universals or laws are descriptive and predictive, and therefore they should not be used interchangeably to refer to regular features of translations. Considering the fact that norms exert influence on translator's strategies, translations of one language pair may always represent certain linguistic features – what Chaim Rabin (1958:144-145) calls 'translation stock' – while those features might not be present in translations belonging to other language pairs. Similarly, Eskola (2004) takes this into account by distinguishing *local translation law* from *universal translation law* (i.e., a *Translation Universal*):

> Consequently, I would rather make a distinction between local and universal translation laws than talk about norms and universals as parallel phenomena. Local laws can be found for example in a certain language pair, text type and time span, whereas universal laws are global tendencies that operate in all translation. The impact of the translation process may result in statistical preferences and characteristics that are distinctive of translating between languages A and B for instance (Eskola 2004:85).

Therefore, we should not confuse features that are specific to a certain language pair (and are the result of norms at play) and those that are inherent universal tendencies. This can be done by juxtaposing findings with those relating to other un- or less investigated language pairs and text-types, which is the purpose of the present study.

2.5 Methodological Issues

In section 2.2, seven limitations in studies on translational Persian were listed. In studies on TUs, there are some methodological issues that should be noted and which are summarized in the following.

An issue highlighted by Teich (2003:22–23), among others, is that Baker's four proposals (1993) are restricted to English as the source or target language. The literature is mostly limited to Western languages, especially English. In order to prove TUs' existence, those studies need to be replicated across a wider variety of languages. Anna Mauranen (2004:65) also suggests that more comprehensive studies are needed, as most of the present works are based on specific language pairs and small corpora. Studying TUs, Sandra Halverson (2013), Bert Cappelle and Rudy Loock (2016), and Teresa Molés-Cases (2019) are the few authors who have appreciated the significance of linguistic typology by addressing the need to consider typological similarities and dissimilarities between source and target languages. The typological nature of the source 'shines through' and is observable in the translation (Teich 2003). Last but not least, Chesterman (2004) provides an account of problems in TU research, ranging from testing and corpus representativeness to operationalization and terminology.

3. Corpus Design and Method

The methodology adopted in this study is explained in three sections: corpus design; corpus markup, annotation and tools; and universal features addressed.

3.1 Corpus Design

In the field of Translation Studies, a comparable corpus consists of two subcorpora of translated and non-translated texts (Baker 1995). The comparable corpus used for this study also consists of two subcorpora: original Persian texts and English-Persian translated texts. On the rationale behind choosing English translated texts, statistics (retrieved on 17 January 2020 from *KhanehKetab*, an online database for publishing statistics) indicate that during the past decade, 22.4% of the books published in Iran are translations, the majority from English to Persian. As informants and regular users of Persian websites, it should be pointed out that Persian website materials are also mostly translated from English into Persian. These driving factors, alongside with the authors' knowledge of English, led us to choose such translated materials. Previous studies have been limited in terms of data size (see 2.3) (for example, 150 paragraphs in Alibabaee and Salehi (2012) or three translations and a comparable number of original texts in a comparable corpus in Esteki et al. (2010)). However, each component or subcorpus of the present study consists of one hundred extracts, each of 3000-word length, taken randomly from books and webpages, thus amounting to 300,000 words for each subcorpus and 600,000 words on the whole. In order to avoid any biased collection and for each sample to be representative of the whole work, each text was divided into three parts (beginning, middle and end) and about 1000 words were randomly extracted from each part (keeping the final sentence complete) to achieve a 3000-word sample. While some samples may be slightly longer or shorter, they are all around 3000 words. In addition, all headings and subheadings were removed, because they have distinctive characteristics and should not be treated as normal texts (e.g., see Khodabandeh 2007; Bluestein 2010). Since some sources, especially webpages, were short and random selection was not possible, several short texts were merged to make them long enough; then the extraction was made accordingly.

As mentioned in section 2, the literature on TUs exhibits some limitations. As some scholars have pointed out (e.g., Mauranen and Kujamäki 2004; Xiao and Hu 2015), although the presence of

similar universals has been reported in non-English translational texts (e.g., Swedish), research in this area has been mostly limited to translational English translated from other European languages. Considering Persian, there are few studies on translational Persian – to our knowledge, all on S-universals comparing one or two translations with their source texts. However, they have failed to move beyond literary texts (see section 2.2 for a list of these limitations). Contrary to the predominance of studies on European languages and small-sized corpus-based works on Persian, all confined to literary texts and books, the present research is based on texts from two non-literary fields in Humanities, psychology and sociology. These expository texts on the topics of Humanities were selected randomly (among others) as an instance of non-literary writings which have been ignored in studies on Persian. If we had selected any other non-literary texts in Humanities, the same approach could be taken in order to find features of translational Persian in uninvestigated text-types. Moreover, the texts were extracted from both general and specialized books and websites on the two topics. For both subject areas, equal numbers of academic materials and popular materials fed into the corpus.

The data for sampling were originally limited to texts published in the last decade. However, in practice we encountered a shortage of books in the two subjects of study, especially in the electronic format, since publications in these two fields are not numerous and most of them are not available digitally. Although there are enough texts available on some subtopics (e.g., motivation, psychology of success), they were not selected because there would have been a lack of variety in data. Therefore, we decided to extend the sampling period to the last 25 years, which helped us get access to a larger number of digitized texts. The comparable corpus is equal in terms of number of samples, size, genre and sampling period; hence the results are directly comparable. Table 1 shows the details of the comparable corpus.

Table 1: Details of the comparable corpus

	Original Persian		**Translated Persian**	
Genres	*Words*	*No. of samples*	*Words*	*No. of samples*
Psychology	151,513	50	150,623	50
Sociology	150,975	50	151,076	50
Total	302,488	100	301,699	100

Additionally, we were faced with serious difficulties in the sampling of books for three reasons. First, many books on topics other than literature, especially in the two specified fields, were not digitally available. Second, books digitally available were scanned books which could be neither edited nor dragged and dropped in the corpus, and a considerable number of them were of low quality and/or not readable by optical character recognition (OCR) software. Third, having performed various pilot tests, practically there is almost no good OCR software for Persian texts, except for MatnYar and Google Drive's OCR tool. These two tools also have their own limitations. For example, a limited input (text) can be entered in MatnYar and Google Drive's OCR tool does not follow the structure of the input text's paragraphs (e.g., it may split a single paragraph or merge two different paragraphs). Moreover, there are sporadic errors and wrong recognitions of words and characters in both. However, we decided to use Google's OCR tool for presenting more accurate output. Consequently, we fed Google's OCR tool with screenshots of the sample chunks in books and put the chunks together to obtain a 3000-word sample of a book. Although it was very time consuming, for the sake of accuracy, we double checked the final samples and compared with their sources to correct any possible mismatches.

3.2 Corpus Markup, Annotation and Tools

Each sample was assigned a header. For books, it provides information about the book title and year of publication; for websites, it includes the title of the text, date of the post and the webpage URL. If the websites' texts were short and sampling was done by assembling two or more pages, then the header would show details of all webpages (title, date of post and URL). In this study, we employed various software, namely Virastyar (normalizer), SeTPer (tokenizer), TagPer (part-of-speech (POS) tagger) and Wordsmith tool version 7 (corpus analyzer). To normalize the data, we employed Virastyar, a Persian MS-Word add-in performing Persian spell checking, character standardization, 'Pinglish' transliteration, punctuation correction and calendar conversion. We also made a list of possible variations in spelling and orthographic forms that might not be covered by Virastyar and double-checked both the subcorpora to ensure a more complete normalization and remove orthographical variations as much as possible (for more information about difficulties and challenges in Persian orthography see Bijankhan et al. 2010; Seraji 2015).

Moreover, for tokenization and POS tagging, we utilized tools developed by Mojgan Seraji (2015) for Persian, namely SeTPer (sentence segmenter and tokenizer) and TagPer (POS tagger). SeTPer has been optimized for Persian as it recognizes Persian punctuation signs like the reverse comma (*virgule*) and angle quotes (*guillemet*) (Seraji 2015:88-90). The POS tagger, TagPer, was developed for Persian using the statistical POS tagger HunPoS and trained on the Uppsala Persian Corpus (UPC), which is a modified version of the Bijankhan corpus (Seraji 2015:91-96). In particular, it benefits from 31 atomic POS tags. The 15 main POS categories are adjective, adverb, clitic, conjunction, delimiter, determiner, foreign word, interjection, symbol, noun, numeral, preposition, preverbal particle, pronoun, and verb (Seraji 2015:92). After training on a subset of UPC, Seraji (ibid.) evaluated the tagger and reported an overall accuracy of 97.46%.

3.3 Universals Under Investigation

Two universal features were selected for investigation, namely simplification and explicitation. There are a number of uninvestigated T-universal candidates in Persian (e.g., explicitation, untypical lexical patterning, under-representation of TL-specific items, etc.). However, these two T-universals were selected not only because prior studies on Persian did not account for them as possible T-universals, but also they are the most discussed TUs in the literature which provide us with more works against which we can compare our results. As mentioned in Section 2.2, simplification may be manifested at lexical, syntactical and/or stylistic levels (cf. Blum-Kulka and Levenston 1983; Laviosa-Braithwaite 1996). Regarding explicitation, although Blum-Kulka's claim that translations tend to exhibit more cohesive markers than non-translations was based on cohesion markers in translations and their corresponding source texts, there are also other studies (e.g., Chen 2006, cited in Xiao 2010) that have addressed translation and native target texts as well.

The presence of universals was identified through a number of features. For simplification, we selected the three signs discussed in Laviosa-Braithwaite (1996) and Xiao and Yue (2009). Laviosa-Braithwaite (1996) designed a comparable corpus of translated and non-translated English texts and finally concluded that translational language uses lower lexical density, shows less lexical variety, and reports greater mean sentence length. Xiao and Yue (2009) also observed that

translated Chinese fiction displays a significantly greater mean sentence length than native Chinese fiction.

For explicitation, we employed the higher frequency of connectives and cohesive ties in translated than non-translated language (Olohan and Baker 2000; Chen 2006). We know that cohesion can be realized through different linguistic features; for instance, it can be lexicalized or established by means of pronouns. Here, conjunctions are examined as cohesive ties. There are mainly two types of conjunctions in Persian: simple (one word) and compound (a combination of two or three words). As native language users of Persian, we selected ten of the most commonly used (Table 2).

Table 2: Ten commonly used cohesive ties in Persian

Cohesive ties	Meaning	Cohesive ties	Meaning
و [*wa*]	and	بلکه [*balke*]	but … also, but … instead
که [*ke*]	that, which	چون [*čown, čūn*]	because, since
اما [*ammâ*]	but, however	زیرا [*zirâ*]	because
ولی [*vali*]	but	بنابراین [banâ bar in]	therefore
اگر [*ágar*]	if	سپس [*sepas*]	then

As can be seen in Table 2, for some Persian connectives meaning is constant while for others it changes based on the context in which they appear. Here, in this study, we aimed at providing their frequency of occurrence to see whether the number of connectives rises in translated language or not.

4. Features of Translational Persian

This section discusses four different lexical and syntactic features of translational Persian in the corpus under investigation, namely lexical density, lexical variety, mean sentence length, and frequency of connectives.

4.1 Lexical Density and Lexical Variety

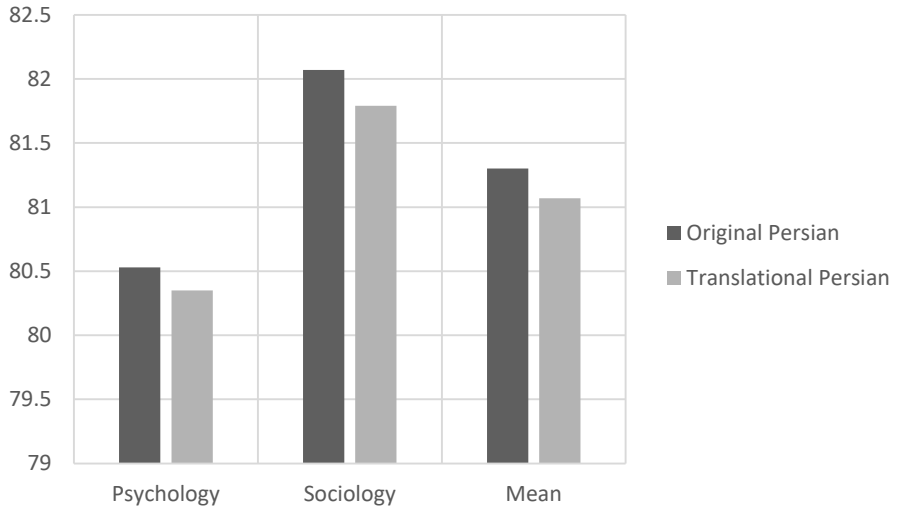
The starting point of our analysis is the hypothesis that translations tend to show a lower lexical density and less lexical variety or lexical richness (see Laviosa-Braithwaite 1996). Sometimes these two terms are used interchangeably, causing terminological confusion. Generally, there are two common approaches to such lexical examinations. One approach is lexical density, defined by Michael Stubbs (1996) as the ratio of lexical items (i.e., content words) to the total number of words. We tagged our corpus on different parts of speech and, therefore, the frequencies of lexical and grammatical POS categories are readily available.

The other approach is TTR, which is the number of unique words (types) divided by the total number of words (tokens). However, the main disadvantage of this quantitative measure is that it is very sensitive to the size of text or corpus and is reliable only when corpora of equal size are compared. Also, when the text reaches a certain length, any increase in new types slows down, which leads to lower values of TTR in larger corpora (for more discussion see Cvrček and Chlumská 2015). To tackle this issue, a newer version was devised by Mike Scott (2004) that is called STTR. It calculates an average TTR of every *n* word (usually consecutive 1000-word chunks) in a given text. It can be concluded that TTR describes lexical richness whereas lexical density measures information load.

Here, at first, we examine the lexical density of translational and non-translational Persian texts in our comparable corpus. Data from several studies suggest a lower lexical density in translational language. For example, Laviosa (1998b) realizes that lexical density is 'highly significantly lower' in translated English (52.87%) compared to non-translated narrative (54.95%). Likewise, regarding non-English languages, Xiao and Yue (2009) find the same trend in translated Chinese (58.69%) and native Chinese fiction (63.19%). Similarly, investigating 15 different genres in the Chinese language, Xiao (2010) notices the same decrease (61.59% in translational vs. 66.93% in native Chinese).

The key question here is whether the same result also holds for translational and native Persian texts. The POS-tagged corpus showed us the lexical and grammatical words. Figure 1 shows lexical density of translational vs. non-translational Persian in the two genres and the mean scores.

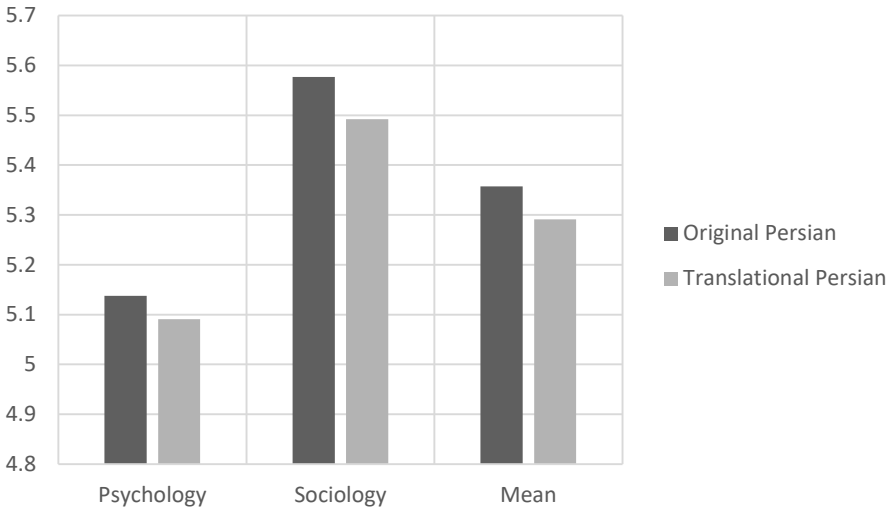
Figure 1: Lexical density in original vs. translational Persian



As the figure shows, original Persian (81.3%) enjoys a higher mean lexical density than translational Persian (81.07%). However, the mean difference -0.23 is not statistically significant (p>0.05). Moreover, sociology texts displayed more lexical density than psychology texts. What stands out in Figure 1 is that although both types of texts show a higher lexical density in native Persian, none of them exhibit a significant difference and it cannot be interpreted as a sign of simplification. These findings are somewhat surprising and contrary to Laviosa's (1998b) observation regarding lexical density in translational English or Xiao and Hu's (2015) findings in translational Chinese. Likewise, regarding the Persian language, information load was higher in translational Persian in studies performed by Esteki et al. (2010) and Alibabaee and Salehi (2012), but it was lower and not significantly different with original Persian in the present study.

Furthermore, Laviosa (1998b) mentions the ratio of lexical words over function words as a measure of information load. Even by calculating this ratio, the result obtained was the same as that of Figure 1. Figure 2 shows that original Persian has a higher ratio of lexical over function words than translational Persian (5.357 vs. 5.291, mean difference=0.066). However, again no significant difference was found in both fields and the mean scores. Comparing the two fields, sociology, as seen in Figure 2, shows a higher ratio of lexical to grammatical words.

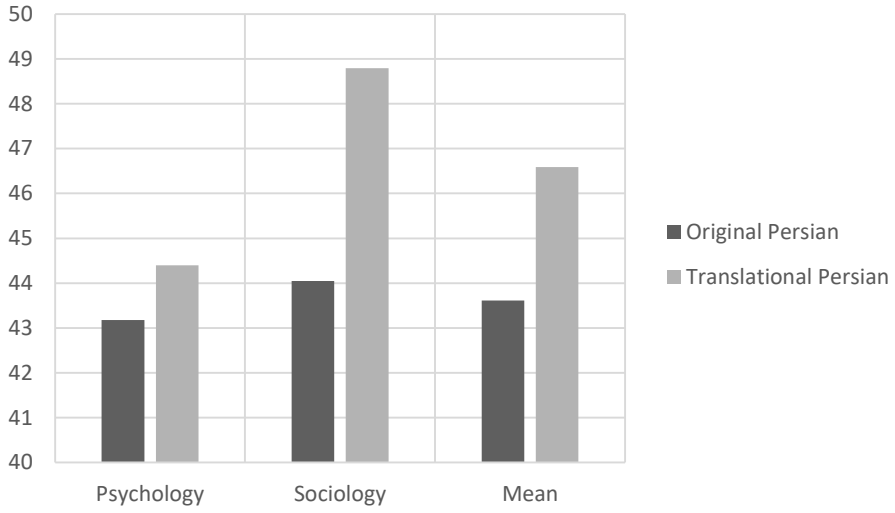
Figure 2: Lexical-to-function word ratios in original and translational Persian



The data in Figures 1 and 2 are similar as both the proportion of lexical words in total number of tokens and the ratio between lexical to function words display the extent of content words of information load. The results do not support, among others, Laviosa's (1998b) hypothesis that the proportion of lexical words over function words is relatively lower in translational English and also differ from Xiao and Yu's (2009) and Xiao's (2010) observations in Translational Chinese.

The second approach of lexical investigation is measuring STTR. In the present study, the STTR was used to remove any minor influence on lexical variety scores caused by corpus size. The findings of statistical analysis using the WordSmith tool indicated two tendencies (Figure 3). Firstly, translations in the comparable corpus showed an unexpected higher mean score in the STTR than non-translations (46.59 vs. 43.61) and the mean difference (+2.98) was statistically significant (t= −5.65 for 198 d.f., p<0.00001). Secondly, this increase in the STTR in translations was reported across both fields. However, the sociology subcorpus showed a higher STTR than the psychology subcorpus.

Figure 3: STTR in original and translational Persian



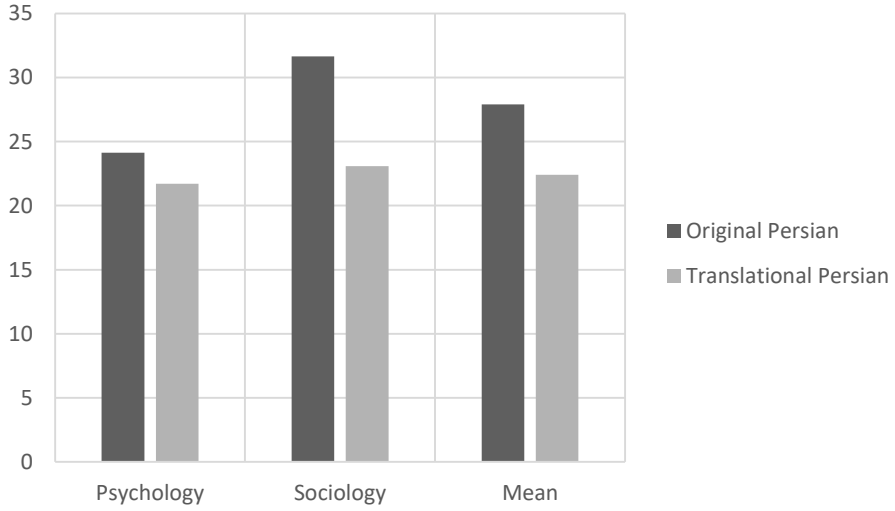
It can be concluded that the vocabulary used in translational Persian is more varied than that of non-translational Persian. This goes against many studies indicating that translations tend to be less varied than non-translations, such as Laviosa-Braithwaite's (1996) observation in English, and Cvrček and Chlumská's (2015) conclusion for Czech, among others. However, considering comparable corpus-based studies on translational Persian, our results of Persian lexical richness confirm Esteki et al.'s (2010) and Alibabaee and Salehi's (2012) findings that a more diverse lexicon is used in translational Persian than in non-translated Persian.

Overall, translational Persian tends to show a lower but not significantly lower lexical density and a higher lexical variety than original Persian texts in the corpus under investigation. These two features are in contrast to earlier findings (Section 2) on translational language tending to be simpler than original writings in terms of lower lexical density and less lexical variety. Consequently, in the current study, lower lexical richness was not supported to be a sign of simplification in Persian translated texts. Hence, instead of considering it as a T-universal in its global sense as stated by Baker (1993), it seems more accurate to consider simplification as a *local translation law* rather that a *universal translation law* (see Section 2.3), or a TU in its narrow sense as argued in Chesterman (2010), present in particular socio-linguistic contexts.

4.2 Mean Sentence Length

The literature has highlighted conflicting findings with regard to mean sentence length as a sign of simplification (see Section 2.2). The analysis of this item, as shown in Figure 4, revealed surprising and counter intuitive results. In the academic environment, especially in Iranian translator training courses, there is a (mis)conception that translations generally exhibit longer sentences than both their sources and native Persian texts — a conclusion that is not compatible with the findings of the present research. Here, the subcorpus of translational Persian texts displayed a significantly lower mean sentence length than their non-translational counterparts (22.4 vs. 27.89), with the difference of -5.49 being statistically significant (t= 6.83 for 198 d.f., p<0.00001).

Figure 4: Mean sentence length in original and translational Persian



A closer inspection of Figure 4 and comparing the texts in the two fields shows that, similarly to results seen in the comparison of lexical density and the STTR, the sociology texts under scrutiny include longer sentences than psychology texts. To put it in another way, it seems that longer sentences is a tendency observed in the sociology texts.

This finding also does not support the previous hypothesis of simplification discussed by Laviosa-Braithwaite (1996) (English), Xiao and Yue (2009) (Chinese), and Esteki (2010) and Mohammad Reza Esfandiari et al. (2012) (Persian) regarding the greater mean sentence length in translational

language being a T-universal feature. The finding, however, is in line with Kirsten Malmkjær's (1997) explanation that stronger punctuations may result in shorter sentences in translated texts. This result is also consistent with Corpas Pastor et al.'s (2008) results reported for Spanish. It is also in accordance with Alibabaee and Salehi's (2012) results, which, albeit working with smaller corpora, suggest that translational Persian tends to exhibit shorter sentences.

4.3 Frequency of Connectives as a Sign of Explicitation

Higher frequency of connectives has been proposed, first by Blum-Kulka (1986), as evidence for explicitation in translations. In this study, as native users of Persian, we selected ten of the most common Persian connectives and scrutinized for any difference in their presence in our comparable corpus of translated and native Persian texts. Table 3 shows the frequency of selected connectives in the two subcorpora.

Table 3: Occurrences of connectives in original and translational Persian

	Psychology Original			**Psychology Translation**			**Sociology Original**			**Sociology Translation**		
Connectives	Frequency	% of running words	Dispersion	Freq.	% of running words	Dis.	Freq.	% of running words	Dis.	Freq.	% of running words	Dis.
و	8,143	5.14	0.97	5,778	3.56	0.94	8,833	5.55	0.96	5,179	2.99	0.97
که	3,957	2.50	0.96	4,688	2.89	0.96	3,656	2.30	0.97	5,203	3.00	0.97
اما	408	0.26	0.89	506	0.31	0.91	401	0.25	0.90	708	0.41	0.95
ولی	183	0.12	0.83	126	0.08	0.79	92	0.06	0.73	152	0.09	0.78
اگر	510	0.32	0.85	436	0.27	0.91	258	0.16	0.88	309	0.18	0.91
بلکه	121	0.08	0.83	125	0.08	0.87	138	0.09	0.81	128	0.07	0.86
چون	109	0.07	0.89	97	0.06	0.72	139	0.09	0.83	175	0.10	0.84
زیرا	97	0.06	0.87	121	0.07	0.80	80	0.05	0.81	31	0.02	0.70
بنابراین	97	0.06	0.83	97	0.06	0.84	94	0.06	0.76	43	0.02	0.75
سپس	30	0.02	0.76	62	0.04	0.79	29	0.02	0.75	37	0.02	0.85
Total	**13,655**	8.63	--	**12,036**	7.42	--	**13,770**	8.63	--	**11,965**	6.9	--

In total, contrary to the hypothesis, there are more connectives in both types of original texts than translational ones (13,655 vs. 12,036 in psychology texts and 13,770 vs. 11,965 in sociology texts). However, there seems to be no clear overall tendency for either subcorpus favoring connectives more than the other. Instead, some connectives were more frequent in translational Persian (که (*kë*), اما (*ammâ*), سپس (*sepas*)) and only one connective (و) was more frequent in original Persian

regarding psychology and sociology texts. Therefore, the findings are contrary to previous studies (among others, Laviosa-Braithwaite (1996) on English, Øverås (1998) on Norwegian, Xiao (2010) on Chinese) which have suggested that explicitation through higher frequency of connectives in translational than non-translational language is a translation universal. On the other hand, the finding corroborates the results of Tiina Puurtinen (2004) who found no clear overall tendency of translated Finnish literature employing connectives more frequently than comparable native texts. The literature on Persian has failed to examine connectives as a possible sign of explicitation by means of comparable corpora and the results of the present study are the first report in this connection.

Another interesting finding was that most connectives followed a disarranging trend (ولی (*vali*) (but), اگر (*agar*) (if), بلکه (*balke*) (but … also, but … instead), چون (*čun, čon*) (because, since), زیرا (*zirâ*) (because), بنابراین (*banâbar-in*) (therefore)) as they were more frequent in one genre but less frequent in the other. These results reflect those of Eskola (2004) and Mauranen (2007) who also suggested that some linguistic features may be genre-bound and language-bound and if we change text-type or the language pair, they might turn out to be examples of a *local translation law* rather than a *universal translation law* (see Section 2.3).

5. Concluding Remarks

This study set out to examine peculiar linguistic features of translational Persian in order to see to what extent the T-universal hypotheses of simplification and explicitation are present in a non-Western language (here Persian). We selected two types of expository from Humanities texts, psychology and sociology, and four features of translational language that can purportedly function as signs of the two T-universals, namely lower lexical density, less lexical variety, greater mean sentence length (all related to simplification), and higher frequency of connectives (connected with explicitation).

First, regarding simplification, the results showed that translational Persian in the corpus under investigation exhibits lower lexical density; however, the mean difference was not statistically significant. Also, the vocabularies in translational Persian represented a significantly more diverse

lexicon than the original Persian. Additionally, regarding mean sentence length, the translational Persian texts in the corpus proved to use shorter sentences than their non-translational counterparts. In all the three features, the sociology texts reported higher scores of lexical density, lexical variety, and mean sentence length than the psychology texts, which can possibly be interpreted as field (also genre, comparing to previous works) variations and their idiosyncratic linguistic features. This interpretation, of course, is worth looking into through recourse to secondary literature. Finally, regarding explicitation, the total number of connectives was higher in the original texts than in translated texts. However, no clear overall tendency was detected in either subcorpus favoring connectives more than the other. Some connectives were more frequent in translations and some in original texts. Further, some connectives followed no trend as they were more frequent in one field but less frequent in the other.

Moreover, the data provide further support for the controversies over the strong version of TU hypotheses (see Sections 2.3 and 2.4). Accordingly, the findings raise intriguing questions regarding the presence of universal features in translations as none of the results for the four features addressed were in line with previously proposed T-universals. Contrary to many previous studies, such as the detailed investigation of Ilisei et al. (2009), features like lower lexical richness and density, greater mean sentence length and higher frequency of connectives might possibly not be among the most salient, *universal* (at least in its global sense) features indicative of the simplification and explicitation hypotheses. Even regarding lexical density, which was lower in translational Persian and followed the same pattern as in the simplification hypothesis, the mean difference was not significant. What the present study reveals may be properties specific to translational (from an English source) and original Persian texts. Taken together, the present study may indicate that, in contrast to what might be assumed, simplification and explicitation as translation universals may not be really universal as discussed by Baker (1993) and Eskola (2004), because they are not universally present in all translated texts, at least as far as translational Persian psychology and sociology texts are taken into consideration. While simplification and explicitation have been supported by a number of studies (see Section 2), they have been challenged by a group of others, especially when language pairs and genres vary and move from the more investigated languages (Western languages, literary texts) to the less investigated ones (non-Western languages, non-literary texts) (see also Chesterman (2004), Mauranen (2007), Xiao and Hu (2015)). It seems

that some of the so-called universal features need to be reclassified under what Eskola (2004) labels *local translation law* rather than *universal translation law* and caution should be applied when any form of such generalizations or 'universal' tendencies are formulated. In this regard, Chesterman (2010:46) criticizes taking such propositions in their absolute universal sense as "a mistake in the first place, a misjudgement about the optimum level of generalization to be aimed at [and] perhaps it would be more fruitful to search for less-than-universal patterns in translation profiles, under different sets of conditions, and thus make more modest claims".

The present investigation differs from most previous research into the linguistic nature of translational Persian in several ways. First of all, considering the fact that the main body of literature is on Western languages, especially English, the Persian language was examined in this research to provide more data than is available in limited studies. Secondly, contrary to much of the literature, here, the authors focused on T-universals and the implementation of comparable corpora. Thirdly, the study carried out was a departure from the more investigated literary genres to less and/or non-investigated non-literary genres (here expository, Humanities texts). A literature review revealed that there is no study on Persian Humanities texts. Finally, the present study tackled most of the limitations listed in Section 2.2, some of them mentioned above. This study included adding variety to the materials (gathering data from both books and online sources) and improving methodological issues. However, this study still had English as the source of translational materials. In addition, it was partially limited by size, since it employed a middle-sized corpus, which might be compensated for in future studies.

Since it is not possible to proceed with any claim about the presence of universal tendencies in translations without validation, further work needs to be done to establish whether TU hypotheses are supported, at least in their current account, in other, especially unexamined, languages and genres. Instead of dismissing the whole possibility of translations displaying common features, we may establish, at least, new tendencies in translations that are different from the previous propositions; for example, simplification in translational language may be universally manifested through features other than lower lexical density or less lexical variety. Moreover, more research is required to account for features of translational Persian and other non-English languages, as well as further research being required for different genres.

References

Ahangar, Abbas Ali, and Seyyedeh Nazanin Rahnemoon (2019) 'The Level of Explicitation of Reference in the Translation of Medical Texts from English into Persian: A Case Study on Basic Histology'. *Lingua, 228,* 102704. https://doi.org/10.1016/j.lingua.2019.06.005

Alibabaee, Ahmad, and Zahra Salehi (2012) 'A Comparative Study of Persian Translated and Non-Translated Political Texts: Focus on Simplification Hypothesis', *Journal of Language, Culture, and Translation*, 1(3): 1-13.

Baker, Mona (1993) 'Corpus Linguistics and Translation Studies: Implications and Applications', in Mona Baker, Gill Francis, and Elena Tognini-Bonelli (eds) *Text and Technology: In Honor of John Sinclair*, Amsterdam: John Benjamins, 233–250.

Baker, Mona (1995) 'Corpora in Translation Studies: An Overview and Some Suggestions for Future Research', *Target*, 7(2): 223–243. https://doi.org/10.1075/target.7.2.03bak

Baker, Mona (1996) 'Corpus-Based Translation Studies: The Challenges that Lie Ahead', in Harold Somers (ed.) *Terminology, LSP and Translation Studies in Language Engineering: In Honor of Juan C. Sager*, Amsterdam/Philadelphia: John Benjamins, 175–186.

Beikian, Ali, Nahid Yarahmadzehi, and Mahta Karimpour Natanzi (2013) 'Explicitation of Conjunctive Relations in Ghabraei's Persian Translation of 'The Kite Runner'', *English Language and Literature Studies*, 3(2): 81–89. https://doi.org/10.5539/ells.v3n2p81

Bijankhan, Mahmood, Javad Sheykhzadegan, Mohammad Bahrani and Masood Ghayoomi (2011) 'Lessons from Building a Persian Written Corpus: Peykare', *Language Resources and Evaluation*, 45(2): 143–164. https://doi.org/10.1007/s10579-010-9132-x

Bluestein, N. Alexandra (2010) 'Unlocking Text Features for Determining Importance in Expository Text: A Strategy for Struggling Readers', *The Reading Teacher*, 63(7): 597-600.

Blum-Kulka, Shoshana (1986) 'Shifts of Cohesion and Coherence in Translation', in Juliane House and Shoshana Blum-Kulka (eds) *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies and Applications*, Tübingen: Gunter Narr, 17–35.

Blum-Kulka, Shoshana and Eddie A. Levenston (1983) 'Universals of Lexical Simplification', in Claus Færch and Gabriele Kasper (eds) *Strategies in Interlanguage Communication*, London: Longman, 119–139.

Cappelle, Bert, and Rudy Loock (2016) 'Typological Differences Shining Through: The Case of Phrasal Verbs in Translated English', in Gert De Sutter, Marie-Aude Lefer, Isabelle Delaere (eds) *Empirical Translation Studies: New Methodological and Theoretical Traditions*, Berlin & New York: Mouton de Gruyter, 235–259.

Chen, Juiching Wallace (2006) *Explication Through the Use of Connectives in Translated Chinese: A Corpusbased Study*, PhD thesis, University of Manchester.

Chesterman, Andrew (2004) 'Beyond the Particular', in Anna Mauranen and Pekka Kujamäki (eds) *Translation Universals: Do They Exist?*, Amsterdam/Philadelphia: John Benjamins, 33–49.

Chesterman, Andrew (2010) 'Why Study Translation Universals?', *Acta Translatologica Helsingiensia (ATH),* 1: 38-48.

Corpas Pastor, Gloria, Ruslan Mitkov, Naveed Afzal and Viktor Pekar (2008) 'Translation Universals: Do They Exist? A Corpus-Based NLP Study of Convergence and Simplification', in *8th Biennial Conference of the Association for Machine Translation in the Americas (AMTA),* 75-81.

Cvrček, Václav, and Lucie Chlumská (2015) 'Simplification in Translated Czech: A New Approach to Type-Token Ratio', *Russian Linguistics*, 39(3): 309–325. https://doi.org/10.1007/s11185-015-9151-8

Esfandiari, Mohammad Reza, Tengku Sepora Tengku Mahadi, Mohammad Jamshid and Forough Rahimi (2012) 'Explicitation and Simplification in Translation of Poetic and Prosaic Genre from Persian into English: the Case of Sadi's *Gulistan*', *Elixir Ling. & Trans*, 44: 7130-7133.

Eskola, Sari (2004) 'Untypical Frequencies in Translated Language: A Corpus-Based Study on a Literary Corpus of Translated and Non-Translated Finnish', in Anna Mauranen and Pekka Kujamäki (eds) *Translation Universals: Do They Exist?,* Amsterdam/Philadelphia: John Benjamins, 83-100.

Esteki, Azadeh, Hossein Vahid Dastjerdi and Hossein Pirnajmuddin (2010) *Testing Simplification Hypothesis: A Corpus-Based Study of Simplification Features in Persian Translated and Untranslated Economic Texts,* M.A. Thesis, University of Isfahan.

Frawley, William (1984) 'Prolegomenon to a Theory of Translation', in William Frawley (ed.) *Translation, Literary, Linguistic and Philosophical Perspectives*, London: Associated University Press, 159-175.

Ghamkhah, Azam and Ali Khazaee Farid (2011) 'A Quantitative Comparison of the Source Text and the Target Text: Three Translations of Jebran's *The Prophet* and a Translation of Al-Ahmad's *Gharb Zadegi*', *Journal of Language and Translation Studies,* 44(3): 103–118. https://doi.org/10.1556/Acr.7.2006.2.2

Gumul, Ewa (2006) 'Explicitation in Simultaneous Interpreting: A Strategy or a by-Product of Language Mediation?', *Across Languages and Cultures*, *7*(2): 171–190. https://doi.org/10.1556/Acr.7.2006.2.2

Hall, Kendra, Brenda Sabey and Michelle McClellan (2005) 'Expository Text Comprehension: Helping Primary-Grade Teachers Use Expository Texts to Full Advantage', *Reading Psychology*, 26(3): 211-234.

Halverson, Sandra (2013) 'Implications of Cognitive Linguistics for Translation Studies', in Ana Rojo and Iraide Ibarretxe-Antunano (eds) *Cognitive Linguistics and Translation: Advances in Some Theoretical Models and Applications*, Berlin: Mouton de Gruyter, 33–74.

Hansen, Silvia and Elke Teich (2001) 'Multi-Layer Analysis of Translation Corpora: Methodological Issues and Practical Implications', in *Proceedings of EUROLAN 2001 workshop on multi-layer corpus-based analysis,* Alexandru Ioan Cuza University: Iasi, 44–55.

House, Juliane (2008) 'Beyond Intervention: Universals in Translation?', *Trans-kom,* 1(1): 6–19.

Ilisei, Iustina, Diana Inkpen, Gloria Corpas Pastor and Ruslan Mitkov (2009) 'Towards Simplification: A Supervised Learning Approach', *Proceedings of Machine Translation*, 25, 21-22.

Kajzer-Wietrzny, Marta (2012) *Interpreting Universals and Interpreting Style*, PhD thesis, Adam Mickiewicz University, Poznań.

KhanehKetab. Available online at [http://www.ketab.ir/search.aspx] (accessed 17 January 2020).

Khodabandeh, Frazaneh (2007) 'A Contrastive Analysis of English and Persian Newspaper Headlines', *Linguistics Journal*, 2(1): 91-127.

Laviosa, Sara (1998a) 'The Corpus-based Approach: A New Paradigm in Translation Studies', *Meta*, 43(4): 474-479. https://doi.org/10.7202/003424ar

Laviosa, Sara (1998b) 'Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose', *Meta*, 43(4): 557–570. https://doi.org/10.7202/003425ar

Laviosa-Braithwaite, Sara (1996) *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*, PhD Thesis, University of Manchester.

Malmkjær, Kirsten (1997) 'Punctuation in Hans Christian Andersen's Stories and their Translations into English', in Fernando Poyatos (ed.) *Nonverbal Communication and Translation: New Perspectives and Challenges in Literature, Interpretation and the Media*, Amsterdam/Philadelphia: John Benjamins, 151–162.

Malmkjær, Kirsten (2007) 'Norms and Nature in Translation Studies', in Gunilla M. Anderman and Margaret Rogers (eds) *Incorporating Corpora: The Linguist and the Translator*, Clevedon: Multilingual Matters, 49-59.

Mauranen , Anna (2004) 'Corpora, Universals and Interference', in Anna Mauranen and Pekka Kujamäki (eds) *Translation Universals: Do They exist?*, John Benjamins, 65-82.

Mauranen, Anna (2007) 'Universal Tendencies in Translation', in Gunilla M. Anderman and Margaret Rogers (eds) *Incorporating Corpora: The Linguist and the Translator*, Clevedon: Multilingual Matters, 32-48.

Mauranen, Anna, and Pekka Kujamäki (eds) (2004) *Translation Universals: Do they exist?*, John Benjamins.

Meyer, Bonnie J.F. and Roy O. Freedle (1984) 'Effects of Discourse Type on Recall', *American Educational Research Journal*, 21(1): 121-143.

Molés-Cases, Teresa (2019) 'Why Typology Matters: A Corpus-Based Study of Explicitation and Implicitation of Manner-of-Motion in Narrative Texts', *Perspectives*, 27(6): 890–907. https://doi.org/10.1080/0907676X.2019.1580754

Morselli, Niccolò (2018) 'Interpreting Universals: A Study of Explicitness in the Intermodal EPTIC Corpus', *inTRAlinea*, Special Issue: New Findings in Corpus-based Interpreting Studies. [http://www.intralinea.org/specials/article/2320]

Olohan, Maeve (2004) *Introducing Corpora in Translation Studies*, Routledge.

Olohan, Maeve, and Mona Baker (2000) 'Reporting That in Translated English. Evidence for Subconscious Processes of Explicitation?', *Across Languages and Cultures*, 1(2): 141–158. https://doi.org/10.1556/Acr.1.2000.2.1

Øverås, Linn (1998) 'In Search of the Third Code: An Investigation of Norms in Literary Translation', *Meta*, 43(4): 557–570. https://doi.org/10.7202/003775ar

Puurtinen, Tiina (2004) 'Explicitation of Clausal Relations: A Corpus-Based Analysis of Clause Connectives in Translated and Non-Translated Finnish Children's Literature', in Anna Mauranen and Pekka Kujamäki (eds) *Translation universals: Do they exist?*, John Benjamins, 165-176.

Rabin, Chaim (1958) 'The Linguistics of Translation', in A. H. Smith (ed.) *Aspects of Translation*, London: Secker & Warburg, 123–145.

Scott, Mike (2004) *WordSmith Tools version 4*, Oxford: Oxford University Press.

Seraji, Mojgan (2015) *Morphosyntactic Corpora and Tools for Persian*. PhD thesis, Uppsala University.

Stubbs, Michael (1996) *Text and Corpus Analysis. Computer-assisted Studies of Language and Culture*, London: Blackwell.

Teich, Elke (2001) 'Towards a Model for the Description of Cross-Linguistic Divergence and Commonality in Translation', in Erich Steiner and Colin Yallop (eds) *Exploring Translation and Multilingual Text Production: Beyond Content*, Berlin: Mouton de Gruyter, 191–227.

Teich, Elke (2003) *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*, Mouton de Gruyter.

Toury, Gideon (1995) *Descriptive Translation Studies and Beyond*, John Benjamins.

Toury, Gideon (2004) 'Probabilistic Explanations in Translation Studies. Welcome as they are, Would they Qualify as Universals?', in Anna Mauranen and Pekka Kujamäki (eds) *Translation Universals: Do They Exist?*, Amsterdam/Philadelphia: John Benjamins, 15–32.

Tymoczko, Maria (2002) 'Computerized Corpora and the Future of Translation Studies', *Meta*, 43(4): 652–660. https://doi.org/10.7202/004515ar

Vahedi Kia, Mehdi and Helen Ouliaeinia (2016) 'Explicitation across Literary Genres: Evidence of a Strategic Device?', *Translation & Interpreting*, 8(2): 82-95.

Xiao, Richard (2010) 'How Different is Translated Chinese from Native Chinese?: A Corpus-based Study of Translation Universals', *International Journal of Corpus Linguistics*, 15(1): 5–35. https://doi.org/10.1075/ijcl.15.1.01xia

Xiao, Richard and Xianyao Hu (2015) *Corpus-based Studies of translational Chinese in English-Chinese translation*, Springer.

Xiao, Richard, and Ming Yue (2009) 'Using Corpora in Translation Studies: The State of the Art', in Paul Baker (ed.) *Contemporary Corpus Linguistics*, London: Continuum, 237–262.

Zasiekin, Serhii (2019) 'Investigating Cognitive and Psycholinguistic Features of Translation Universals' *Psycholinguistics*, 26(2): 114–134. https://doi.org/10.31470/2309-1797-2019-26-2-114-134