Specialized Translation Teaching Strategies: A Corpus-Based Approach

Yingying Ding

Abstract—This study presents a methodology of specialized translation with the objective of helping teachers to improve the strategies in teaching translation. In order to allow students to acquire skills to translate specialized texts, they need to become familiar with the semantic and syntactic features of source texts and target texts. The aim of our study is to use a corpus-based approach in the teaching of specialized translation between Chinese and Italian. This study proposes to construct a specialized Chinese - Italian comparable corpus that consists of 50 economic contracts from the domain of food. With the help of AntConc, we propose to compile a comparable corpus in for translation teaching purposes. This paper attempts to provide insight into how teachers could benefit from comparable corpus in the teaching of specialized translation from Italian into Chinese and through some examples of passive sentences how students could learn to apply different strategies for translating appropriately the voice.

Keywords—Corpus-based approach, translation teaching, specialized translation.

I. INTRODUCTION

CORPUS linguistics is seen as the study of linguistic phenomena through large collections of texts. In recent years, the corpus linguistics and its applications present an important area in translation studies. In 1993 Mona Baker [2] introduced the term Corpus-based Translation Studies, which refers to the study of the process and product of translation through corpora. According to [8], this field of research has the following characteristics. Firstly, the research is based on the analysis of authentic texts. Secondly, this research offers analytical perspectives both in quantitative and qualitative terms so that statistical data can be obtained concerning lexical, syntactic and stylistic characteristics. Finally, this methodology can be applied through linguistic and cultural approaches in translation studies. The primary aim of this work is to compile a small specialized corpus of contracts in the field of food.

This paper is divided into five parts. After a brief introduction on the discipline of Corpus-based Translation Studies, we review the state of the art of the didactic value of corpora in translation teaching. Then a description of the corpus design and compilation will be given, and finally after the analysis of the active and passive sentences some possible applications of corpus in specialized translation teaching will be discussed.

Yingying Ding is with the Department of Interpreting and Translation, University of Bologna, Forlì, 47121, Italy (e-mail: yingying.ding2@unibo.it).

II. CORPUS TYPOLOGY IN CORPUS-BASED TRANSLATION STUDIES

Depending on the type of task to be performed, different types of corpora may be classified. Generally, corpora can be applied to translation studies include parallel corpus and comparable corpus as in [9]. A parallel corpus, as [12] suggests, includes the texts of the source language with their translations. Parallel corpora can be unidirectional, containing original texts in language A (e.g. Italian) and relative translations in target language B (e.g. Chinese), or bidirectional, containing original texts in two languages A and B (e.g. Italian and Chinese) and corresponding translations in the same two languages (e.g. from Italian to Chinese and from Chinese to Italian). Parallel corpora can be applied in translation to identify linguistic equivalents, in the study of translation strategies adopted by translators and in the development of automatic translation systems [12].

A comparable corpus is divided into comparable bilingual corpora and comparable monolingual corpora. The first is defined as a set of original texts in language A and in language B, containing texts with similar content, domain or communication function; the second one consists of original texts in language A and translation in the same language. The comparable corpora have been used mainly for terminology and assisted translation. They can also provide references to the language in its natural context and thus allow translation students and translators to produce high quality target texts [13].

III. CORPORA IN TRANSLATION TEACHING

Traditionally, in translation classes, the teacher plays a central role; s/he should need to be familiar with translating different texts types and genres and to be prepared to comment on student's translation. On the other hand, the students have to acquire language skills and translation ability to "reproduce what is conveyed in the source text by using appropriate translation skills" [7].

For a long time, the translation teaching has been carried out mainly based on the teacher's experience. However, the use of corpora in translation teaching has attracted a growing interest in the last twenty years.

Bernardini [3] proposes to use large corpus in the teaching activity which is helpful to the development of students' professional skills such as "*awareness, reflection and reaction*". Aston [1] shows the main benefits deriving from the use of corpora in terms of training and education, as corpora provide a tool to facilitate the translation process, giving translators the opportunity to expand their own linguistic and

World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:12, No:5, 2018

cultural horizons. Reference [16] reveals the value of comparable corpora in translation teaching, arguing that the use of corpora and concordance software can allow students to understand the languages and cultures involved and to develop their awareness of the relationships between possible equivalents. Also Cai [5] confirms the use of parallel corpora in the teaching of translation, particularly concerning collocations and phraseology.

Observing the behavior of linguistic characteristics in parallel bilingual corpora enables students to make better choices about tasks and texts types. The author finally suggests to use corpora in order to elaborate didactic materials focused on lexical elements and typical structures (idiomatic expressions, collocations).

IV. CORPUS DESIGN AND COMPILATION

As already demonstrated in many studies, the comparable corpora can provide a valuable tool for translation teaching. Reference [12] points out that the comparable corpora can be useful "to identify text-type-specific formulations, validate intuitions and provide explanations for appropriateness of certain solutions to problems" in translation. In this regard, a comparable corpus created with well-defined objectives and criteria allows students to deepen their knowledge of language for specific purposes. Depending on the purpose of the study, the creation of comparable corpora intended for teaching and learning of translation must take into account different criteria, including dimension, domain, and especially textual typology and genre [13]. In the following section it will be described the most relevant criteria adopted in this paper for the construction of comparable corpora.

A corpus consists of a large collection of authentic texts and represents a variety of language. One of the first criteria to be addressed in the corpus design phase is therefore the dimension. However, there are no well-defined rules that can establish the maximum or minimum dimensions of a corpus, which vary according to the objectives for which the corpus was designed. According to Sinclair [15], a general corpus should be "as large as possible, and should keep on growing". If we intend to conduct studies on every linguistic level and in any communication field, it is natural that the corpus should be as large as possible and represent different textual and linguistic varieties. Compared to a general corpus, a specialized corpus is representative of a particular variety of language or of a restricted domain, so the dimension can therefore be smaller. As stated by [14] the specialized corpora "can be smaller than those used for LGP studies" (language for general purposes). Reference [4], on the other hand, argues that the size may not be an important condition because the adequacy of a corpus also depends on its application. For this reason, it is essential to evaluate the collected data to have a high quality specialized corpus that meets the specific needs for which it was complied. Based on these reflections, it has been chosen to leave the corpora flexible so as to correct and expand them during the course of the work, making them more specific for the didactic purposes. The initial DIY corpora created for the present study consists of a sub-corpus in Italian and a sub-corpus in Chinese

(of about 63,000 words).

In translation, the text is a communicative unit through which the translator interacts with the source text in another language. In the data collection phase, the genre that we decide to investigate is contract text. All texts were collected from websites and converted to text format .txt with the AntFile Converter software [17]. The Notepad ++ program was used not only during transcription, but also for the revision and manual correction of the texts [18]. In order to analyze the corpus, we chose to use the AntConc software, version Windows (3.4.3) [19], which was designed by Laurence Anthony. After collecting the texts, it was necessary to address another important passage related in particular to the texts in the Chinese language, the tokenization.

In many texts, the boundaries of distinct words are generally denoted by a space. However, some languages, like Chinese, do not use space to separate words. As a non-segmented language, the tokenization in this case is fundamental to divide the sequences of Chinese characters into minimal units called "tokens" - words, punctuation, dates, numbers, abbreviations and other - in order to analyze texts [11].
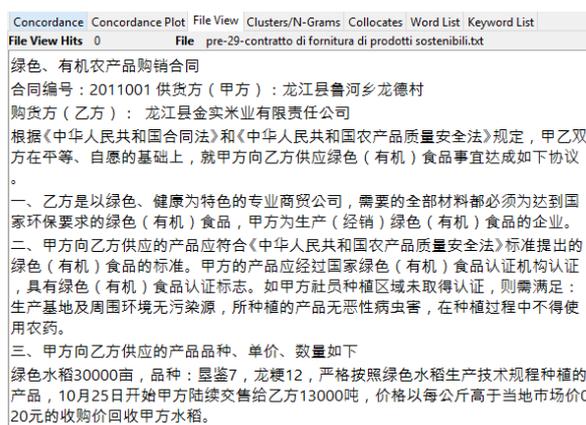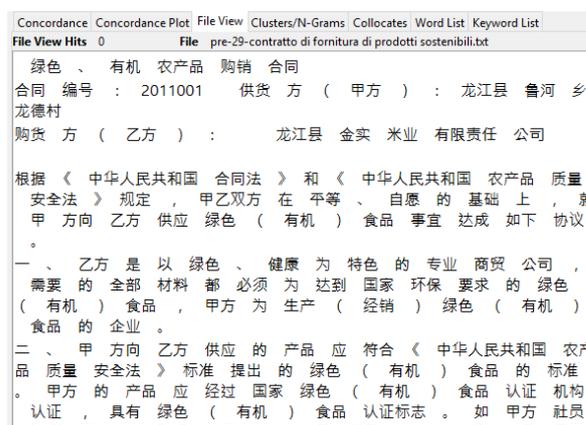

Fig. 1 Non-segmented corpus

Fig. 2 Segmented corpus

The tool used for this study is SegmentAnt (Version 1.1.3), created by Laurence Anthony to segment Japanese and Chinese, which can be downloaded free from the web [20].

SegmentAnt was designed on the basis of various Chinese segmentation tools such as Jieba, PyNLPIR, Tinysegmenter, and Smallseg. With this tool you can select a text or a series of texts (UTF-8 encoded) and segment them by separating words. Here are two examples of segmented text and non-segmented text (see Figs. 1 and 2).

As we have seen, the word segmentation is a necessary step for analyze Chinese texts. In the section that follows we will explore some features of passive constructions in Italian and Chinese on the basis of comparable corpus data.

V. COMPARATIVE ANALYSIS: ACTIVE AND PASSIVE SENTENCES

The contract text is a formal written agreement, which creates legal obligations between two or more parties. When we translate contracts, since Chinese and Italian belong to two different language families and present many differences in their grammar and discourse structure, in order to translate active and passive sentences, further consideration of the strategies and techniques will also be required.

A. Passives in Italian

The passive in Italian is generally marked by the auxiliary *essere* 'be' followed by a past participle (see example 1). The structure 'essere + past particle' can be considered as the norm for Italian passive, which focuses on the action or on the person/thing affected. However, *essere* can also be replaced by other auxiliary such as *venire* 'come' and *andare* 'go' (see examples 2 and 3).[1] In the example 2, the passive structure is formed with *venire* and the past participate. This is used only in a more formal register and tends to express the idea that a regular action is involved. In the example 3, the structure *andare* is followed immediately by the past participle, in which case it has a prescriptive sense, indicating how things should be done.

(1) I prezzi dei prodotti forniti **sono indicati** nel listino allegato[...].
   'The prices of the products supplied are indicated in the attached price list'.

(2) Il contratto **viene eseguito** secondo le seguenti modalità.
   'The contract is executed according to the following procedures'.

(3) [...] **va considerato** non essenziale per la Fornitrice.
   'it should be considered not essential for the Supplier'.

Comparing the passive forms in the contract contest, the frequencies are given in Table I.

---

[1] Another way of expressing the passive is the construction with si, especially in the more colloquial register. When the verb used is an impersonal *si* construction has an expressed subject, it can express a passive meaning. For example, it is said that *In quel negozio si vendono delle belle scarpe* 'they sell some nice shoes in that store' is equivalent to *In quel negozio sono vendute delle belle scarpe* 'some nice shoes are sold in that store' [6]. The impersonal sentences in Italian are so similar with the *si* passive construction. Considering this feature, it cannot be studied efficiently using a corpus-based approach, so we will not include such type of passive voice in this study.

TABLE I
FREQUENCIES OF PASSIVES IN CORPUS IT

Corpus IT	Essere passive	Venire passive	Andare passive
frequency	217	57	5
%	77.8%	20.4%	1.8%

It can be seen from Table I that *essere* passives are predominantly more frequent than *venire* passives and *andare* passives, especially the last one, which is generally used in colloquial speech. Since the passive voice is often used to highlight the patient and its consequences, to evaluate this hypothesis we examined all passive sentences (279 instances). The results are given in Table II.

TABLE II
SEMANTIC PROPERTIES OF PASSIVES

Passive type	Negative (%)	Positive (%)	Neutral (%)
Essere passive	3.2%	3.7%	93.1%
Venire passive	1,8%	7%	91.2%
Andare passive	——	——	100%

From Table II, it is clear that the distribution of these passive constructions is grouped into three categories. We note that the passive sentences are more frequently neutral in this genre. The portions of *essere* passives and *venire* passives are very similar. It is also important to note that the distribution of semantic properties is possibly related to the contract texts, which are supposed to be precise, clear and neutral. In the next section, we will analyze the passive constructions in Chinese and compare them with Italian passives.

B. Passives in Chinese

Passive structures in Chinese are often marked by the passive particles, such as *bei*, *shou* and *you*, as shown in (4), (5) and (6). In general, *bei* is a common passive marker in Chinese, which can mark passive structures with or without an agent. In contrast to *bei*, which is a function word, *shou* 'be subjected' is a lexical verb with passive meaning. However, *you* as a passive marker expresses obligations, duties or responsibilities.

(4) 被　执法　部门　处罚
   'Punished by law enforcement authorities'

(5) 受 不可抗力　影响　的　一方
   'the Party affected by force majeure event'

(6) 由　双方　代表　签字
   'Signed by both parties'

Table III gives the frequencies of passive constructions with three different markers in the Chinese corpus. It can be seen that the construction with *you* appears to be the most frequent in our data. Although the sentence with *bei* is the most commonly used in Chinese, however, the statistics suggest that in terms of this specific genre, the frequency of passive markers varies greatly. Furthermore, the original meaning of the participle *bei* is 'suffer', consequently, when we use it, the situation described by the *bei* construction is often interpreted negatively, same as *shou*. Reference [10] points out, however, that although the use of the passive constructions to express a

negative evaluation is still common, they have spread to neutral contexts as well, especially in written Chinese.

TABLE III
FREQUENCIES OF PASSIVES IN CORPUS ZH

Corpus ZH	*bei*	*shou*	*you*
frequency	4	5	92
%	3.96%	4.95%	91.09%

TABLE IV
SEMANTIC PROPERTIES OF CHINESE PASSIVE MAKERS

Passive type	Negative (%)	Positive (%)	Neutral (%)
bei	50%	——	50%
shou	100%	——	——
you	40.2%	——	59.8%

Table IV shows the distribution of passive markers across three meaning categories. We can note that Chinese passives usually have negative semantic property. The passives marked by *bei* and *shou* are always negative because of their original meaning. In relation to *you*, it can be seen that it shows a less percent of negative property, because *you* does not have an inflicting meaning. Considering these observations, the passive with *you* can be used more frequently than *bei* and *shou* for contract texts.

VI. CONCLUSION

In this study we have attempted to show how comparable corpora can be used to investigate passive voice of translation. We started out by reviewing the existing literature on corpus-based translation studies, especially on translation teaching. We described the creation of a small comparable corpus for the specialized translation teaching. In our analysis of passive constructions in Italian and Chinese from a contrastive perspective based on the comparable corpus, we noted that there are some differences in terms of frequencies and semantic properties. These differences are primarily associated with the functions of passive markers and the genre involved in this study. Given the small size of the comparable corpus complied for this paper, it is difficult to provide an abundance of empirical evidence. However, it is hoped that comparable corpora could offer a useful tool for specialized translation teaching.

REFERENCES

[1] Aston Guy (2000). "I corpora come risorse per la traduzione e per l'apprendimento", in *I corpora nella didattica della traduzione*. *Corpus Use and Learning to Translate*. Ed. S. Bernardini e F. Zanettin, Bologna: CLUEB, pp. 21-29.
[2] Baker, Mona (1993). "Corpus Linguistics and Translation Studies: Implications and Applications". In *Text and Technology: In Honour of John Sinclair*, ed. Mona Baker, Gill Francis & Elena Tognini-Bonelli. Amsterdam & Philadelphia: John Benjamins. Pp. 233-250
[3] Bernardini, Silvia (2000). "I corpora nella didattica della traduzione: dall'addestramento alla formazione". in *I corpora nella didattica della traduzione*. *Corpus Use and Learning to Translate*. Ed. S. Bernardini e F. Zanettin, Bologna: CLUEB, pp. 81-103.
[4] Biber, Douglas (1993). ''Representativeness in corpus design.'' *Literary and Linguistic Computing*, 8(4). Oxford: Oxford University Press, pp. 243–257.
[5] Cai, Yonggui (2015). *Binlingual Paralle Corpus: Based vocabulary Translation Teaching*. Journal of Yunnan Agricultural University, 2015,

9 (3): 103-107
[6] D'Alessandro, Roberta (2007). *Impersonal "si" Constructions*. Berlin, Mouton de Gruyter
[7] Hu, Kaibao (2011). *Introducing Corpus-based Translation Studies*. Shanghai: Shanghai Jiao Tong University Press.
[8] Laviosa-Braithwaite, Sara (1998). "Universals of Translation." Mona Baker (eds): *Routledge Encyclopedia of Translation*. London: Routledge, pp. 288-291.
[9] Laviosa, Sara (2002). *Corpus-based translation studies. Theory, findings, applications.* Amsterdam: Rodopi.
[10] Li Linding (1986). *Xiandai hanyu juxing.* (Sentence patterns in modern Mandarin.) Peking: Shangwuyinshuguan.
[11] Jurafsky, Daniel & James H. Martin (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River NJ: Prentice Hall.
[12] Olohan, Maeve (2004). *Introducing Corpora in Translation Studies*. London/New York: Routledge.
[13] Pearson, Jennifer (1998). *Terms in Context*. Amsterdam/Philadelphia: John Benjamins.
[14] Pearson, Jennifer (2003). "Using Parallel Texts in the Translator Training Environment", in *Corpora in Translator Education*, ed. Federico Zanettin, Silvia Bernardini and Dominic Stewart, Manchester: St. Jerome Publishing, pp. 15-24.
[15] Sinclair, John (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
[16] Zanettin, Federico. (2001) "Swimming in words: Corpora, translation, and language learning", in *Learning with Corpora*. In G. Aston, Houston, TX: Athelstan, pp. 177-197.
[17] http://www.laurenceanthony.net/software/antfileconverter/ Accessed on 10/03/2018
[18] https://notepad.it.softonic.com/ Accessed on 10/03/2018
[19] http://www.laurenceanthony.net/software/antconc/ Accessed on 10/03/2018
[20] http://www.laurenceanthony.net/software/segmentant/ Accessed on 11/03/2018