

Sketching the changing patterns in kaleidoscopes: New developments in corpus-based studies of translation features (2001–2021)

Shuangzi Pang – Kefei Wang
Shanghai Jiao Tong University / China
Beijing Foreign Studies University / China

Abstract – Corpus-based Translation Studies (CTS) have developed and advanced substantially since its emergence in the 1990s. This article provides an overview of the evolution of CTS from 2001 to 2021, identifying new challenges and research opportunities. The evolution of CTS is presented into two stages: the establishment of the subject matter and the expansion of research, respectively. We argue that CTS may enter the stage when the traditional specialties, such as using corpora in contrastive linguistics and translation, continue to advance, while a variety of new research points emerge and expand. After outlining current problems and unresolved issues, the analysis presents newly emerged research areas, assumptions, perspectives, and cross-fertilization with neighboring disciplines as the new developments in CTS. Four possible trends in CTS are framed and presented accordingly. The analysis highlights the significant advancements made in CTS over the past two decades and provides a valuable resource for researchers and practitioners interested in understanding the current state of CTS, and the directions it may take in the future.

Keywords – corpus-based translation studies; the third code; translation features; expansion; socio-cognitive constructs; interdisciplinary

1. INTRODUCTION¹

Corpus-based Translation Studies (henceforth CTS) have developed and advanced substantially since they emerged in the 1990s. As Kenny (2001: Introduction) states:

Kaleidoscopes allow us to view patterns, and to change those patterns at will. In corpus linguistics, the words and characters of electronic texts act like pieces of coloured glass and paper, constantly forming new patterns, which then recede as others take their place.

¹ This research has been funded by *The National Social Science Fund of China* (20BYY020). The grant was awarded to the project *Influence of Language Contact through Translation on the Register Features of Vernacular Chinese after May Fourth Movement in China*. We would also like to express our sincere gratitude to Sara Laviosa for her feedback and support throughout the course of the research.



Over the past three decades, the kaleidoscope of CTS has continuously witnessed the emergence and evolution of new patterns, driven by the creation and utilization of novel types of corpora. These corpus-based approaches have yielded valuable insights into translation theories and practices, encompassing diverse languages and disciplines. The notion of the ‘third code’ —the fact that “translation is essentially a third code which arises out of the bilateral consideration of the matrix and target codes” (Frawley 1984: 257)— has been extensively studied as an independent variety, shedding light on the underlying motivations behind this particular phenomenon. Moreover, the exploration of translation universals has led to the creation and investigation of new technical terms within this domain, such as ‘interpretese’ (Kajzer-Wietrzny 2018: 93), ‘foreignese’ (Kajzer-Wietrzny 2018: 93), and ‘varioversals’ (Szmrecsanyi and Kortmann 2009: 33). Through the lens of corpora, previous studies have successfully explored the universality and diversity of translation, reflecting the evolving perception of this field. However, the research focus concerning the features found in translated texts has expanded beyond the mere verification or falsification of alleged translation universals. It is now established that these alleged universals are influenced by various factors, including register, interference from the source language, social norms, socio-contextual factors, language contact, and cognition (Halverson 2003; Kruger 2018a). As of today, CTS has evolved into a distinct field, incorporating increasingly sophisticated methodologies that draw from and contribute to contrastive linguistics, translation features, translation norms, translators’ style, language contact, discourse analysis, and cognitive translation studies. Notably, the study of the third code has expanded its scope and depth, encompassing a broader range of observations.

This article presents a global picture of the recent trends and cutting-edge avenues CTS has experienced in the last two decades in the domain of translation features. Section 2 provides an overview of what CTS has experienced and a range of topics and methodologies which have been used in this area for the past two decades. The evolution of CTS is divided into two stages —divergence and convergence— due to the extent of its dependency on other disciplines. Section 3 discusses some unresolved issues and current problems with CTS. Section 4 concentrates on the newly emerging areas, perspectives, and assumptions in CTS in recent years. Finally, four possible trends are framed accordingly in Section 5 and some final remarks are provided in Section 6.

2. DIVERGENCE AND CONVERGENCE: WHAT CTS HAS EXPERIENCED IN THE PAST DECADES

By the end of the twentieth century, the notion of ‘translation universals’ had gained prominence following the publication of Baker (1993). In the previous decades, the study of translation universals had produced a fruitful line of research in CTS, which may be divided into two stages.

2.1. Stage I: Establishment of the traditional research object in CTS

According to Shneider (2009), the evolution of a scientific discipline can normally be divided into four stages. Chen (2017: 5) further specifies that the object of the research is established in the first stage. The second stage is characterized by the development of research instruments, or tools, which enable researchers to investigate underlying phenomena. The third stage is a prolific stage, because many results are produced, and the understanding of the research problems is substantially advanced. Finally, in the fourth stage, the original specialty may continue to be investigated along the original research agenda, with the tools developed by the original specialty contributing to the development of other subject domains (Chen 2017: 5).

Applying the division mentioned above to the development of CTS, the scrutinization of the research object (i.e., translation universals) has had parallels with the development of electronic texts and text analysis tools in CTS. This can be exemplified by the fact that the *Translational English Corpus* (TEC)² was developed in the mid-1990s when Baker (1993) brought together two strands, namely Descriptive Translation Studies (DTS) and Corpus Linguistics (CL) and proposed the notion of translation universals. In other words, the evolution of CTS cannot be separated from technological tools in the beginning, as evidenced by what Zanettin (2012: 12) argues:

a comparison of linguistic data in comparable monolingual corpus of translated and non-translated texts could unveil some regular patterns of behavior common to all translated texts.

With the above in mind, the evolution of CTS can be categorized into two stages. In the first stage (1993 to 2010), the research focus was established, and numerous studies emerged to either verify or challenge it. During this period, CTS aimed to set itself apart from other disciplines (De Sutter and Lefer 2019: 55), granting translated texts their

² <https://genealogiesofknowledge.net/translational-english-corpus-tec/>

own rightful place. Concurrently, corpora (such as TEC), research instruments, and tools were developed to investigate underlying phenomena. Subsequently, several parallel and comparable corpora, such as the *English-Norwegian Parallel Corpus* (ENCP)³ and the *English-Portuguese Bi-directional Corpus* (COMPARA),⁴ were compiled to explore various aspects and advance research in the field.

As De Sutter and Lefer (2019) argue, in its early years, CTS strongly focused on delineating the unique characteristics of translation in order to distinguish it from other disciplines. From 1993 to 2010, the primary areas of investigation included verification and falsification of translational features, contrastive studies between different languages, translators' style, and translation norms, among others. Building upon the assumption of translation universals, Klaudy and Károly (2005) proposed the 'asymmetry hypothesis', which assuming that the notions of 'explicitation' and 'implication' are not symmetric translation strategies, refined and expanded upon the existing 'explicitation hypothesis' by considering factors such as implication and translation direction. This hypothesis aimed to bridge gaps in previous studies and provide a comprehensive framework. During this stage, the research scope gradually solidified, and a significant number of studies were conducted to examine the translation features under scrutiny. Numerous scholars conducted innovative studies on translation features from an English-Chinese perspective during this stage. For example, Wang and Qin (2009) analyzed the general features of translated Chinese by comparing translated and non-translated Chinese texts. Their study revealed that translated Chinese exhibited a higher type-token ratio and employed longer sentence segments. However, their results did not fully support the hypothesis of translation universals. Additionally, Xiao (2010) explored the potential lexical and syntactic features of translational Chinese using two comparable corpora —the *ZJU Corpus of Translational Chinese* (ZCTC)⁵ and the *Lancaster Corpus of Mandarin Chinese* (LCMC)⁶— and found that translated Chinese utilizes conjunctions more frequently, suggesting a tendency toward explicitation.

³ <https://varieng.helsinki.fi/CoRD/corpora/ENPC/>

⁴ <https://lindat.mff.cuni.cz/repository/xmlui/handle/11372/LRT-866>

⁵ <https://www.lancaster.ac.uk/fass/projects/corpus/ZCTC/>

⁶ <https://www.lancaster.ac.uk/fass/projects/corpus/LCMC/>

2.2. Stage II: Expansion of the research issues

The second stage in Shneider's (2009) model ranges from the timeframe when Becher's (2010) influential work was published to the present time, when CTS intends to expand and converge with other neighboring subjects (see Section 2.1 for the first stage). During this stage, the existence of a third code has become more doubtful, and its study has been substantially expanded.

As a breakthrough in CTS, which gained no less attention than Baker's (1993) work, Becher (2010: 1) claims that the dogma of translation-inherent features rests on fallacious theoretical considerations and premature interpretations of empirical data. According to a science mapping (see Section 4), it has been proven that Becher (2010) becomes the most cited reference in the literature on the topic during the following decade in CTS, which indicates that scholars realize that the hypothesis of translation universals is rather problematic. In 2010, there was a consensus that certain factors, such as source language interference and conservatism, were not adequately addressed in Olohan and Bake (2010) on the analysis of the translation of *that*.

Kruger (2012: 335) innovatively shows that the recurrent features, or universals, of translated language are primarily the result of a "mediation process" rather than the particularities of bilingual language processing. More recently, as De Sutter and Kruger (2018: 55) point out, most studies have witnessed a process of removing the interdisciplinary walls. In other words, CTS is being converged with neighboring disciplines, such as interpreting studies, contrastive linguistics, second language acquisition, variational linguistics, sociology, psycholinguistics, and contact linguistics, among others. Studies dealing with the third code have also concomitantly profited from a variety of new research agendas, paradigms, and settings.

As far as a paradigm is concerned, the use of solely parallel or comparable corpora has been evolving in combination. The emerging composite corpus paradigm, which combines both parallel and comparable corpus, has shown to be an innovative approach to studying the relationship between translation and language change. Through the kaleidoscope of composite corpora, we can see source language shining-through effects and the effects of translated languages on the target languages. The indirect influence of English-Chinese translation on the original in Chinese can thus be empirically measured on the basis of the new type of corpus, which incorporates a parallel and comparable corpus.

Regarding new settings, specialties —such as studies of translational features and contrastive studies between languages— have been intersected with the areas of education, language contact, translation evaluation, translator style, translators' expertise, editor's invention, legal translation, news translation and medical translation, etc. These have provided strong evidence for the notion that translated language is a third code shaped by the socio-cognitive constraints that operates in mediating process.

In this study it is proposed that CTS may enter the stage when traditional specialties —such as using corpora in contrastive linguistics and translation— continue to advance, while a variety of new research agendas emerge and are expanded at the same time.

3. CURRENT PROBLEMS IN CORPUS-BASED TRANSLATION STUDIES

Although significant developments have been made in the field of CTS in recent years, there are still unresolved issues. Kruger (2018b) identifies three blind spots in CTS.

The first blind spot is the underestimation of the complexity of translated features. As Kruger (2018a: 9) claims, the hypothesis of translation universals downplays the complexity of translation, which has proven to be constrained by register, source language interference, translators' expertise, translation methods, translation direction, cross-linguistic differences, and socio-cognitive factors in recent years. For example, Redelinghuys and Kruger (2015) agree with the hypothesis that linguistic operationalizations of the translated features demonstrate significant differences in the work of experienced and inexperienced translators. Their study unveils that the translators' expertise may constrain translated features.

The second blind spot concerns the vagueness of conditioning forces, which are difficult to pinpoint. The conclusions yielded in different studies (Øverås 1998; Olohan and Baker 2000; Dimitrova 2005) are difficult to compare and replicate because the indicators are selected and operated differently. According to Kruger (2018b), what is challenging is the disentanglement of the various explanatory hypotheses proposed for the features of translated language. Currently, the causes for translation-inherent features are normally attributed to cross-linguistic influence from source languages, risk-avoidance, conservative overadjustment to the norms for formal writing, and cognitive complexity, such as chunking and entrenchment, which underlie these

differences. De Sutter and Kruger (2018) have adopted a multifactorial approach to address this issue in a corpus study. The study aims at separating the three proposed explanations for explicitation, leading ultimately to a better understanding of what translation is, how it is affected by different circumstances, and how it relates to other types of constrained communication.

Even though the existing studies pointed out above have provided an innovative and in-depth perspective of the forces motivating translation universals, what has garnered great popularity is still confined to synchronic research. There is a paucity of attention given to the investigations from a diachronic dimension. Although Hermans (1999: 155) has been keen to raise the importance of ideology and socio-historical context, previous studies have been unable to model social causation and too much emphasis is put on stability rather than on change. In the Chinese scholarship, Ke (2005) also argues that explicitation and implicitation are the key issues not only discussed in linguistic levels but also embedded in socio-historical contexts.

The third blind spot is related to the fact that, although there is currently a new consensus on the combination of theoretical frameworks and CTS research paradigms, the methods on how to integrate them are seldom investigated. Most studies have explored the assumption of translation universals, but how to combine CTS and other disciplines to gain a profound understanding of the nature of translation needs further exploration.

Methodologically, previous studies can be framed as three strands: 1) traditional research based on parallel corpora; 2) new large-scale quantitative research based on comparable corpora; and 3) composite research based on a combination of parallel and comparable corpora. There is a deficiency of internally coherent multi-lingual composite corpora incorporating parallel and comparable corpus covering different registers. While composite corpora have been employed in some studies (Bisiada 2013; Malamatidou 2018), knowing how to exploit them best remains a challenge. It is also difficult to annotate a large corpus with accuracy and precision on semantic, syntactic, and pragmatic levels, and spoken corpora are even more challenging to compile due to difficulties in data collection and transcription. Furthermore, although more complicated predictive models are designed, how to utilize them to tackle key issues in CTS remains a challenge. Particularly, using corpora properly and effectively to detect conceptual problems is still one of the difficulties in translation studies.

4. NEW DEVELOPMENTS IN CTS IN RECENT YEARS

In this section, we present new research issues and the progress of methodologies in recent years regarding CTS. New research issues will partly be based on a visual analysis of science mapping. To achieve a global picture of the new development of CTS, we visualize and analyze a dataset from *Web of Science* (WOS) ranging from 2001 to 2021. To do so, we make use of a new version of *CiteSpace 5.0*.⁷ The input data of our review is generated by the results from search queries in WOS. WOS is an interdisciplinary database with records from several bibliographic databases, among them *Science Citation Index Expanded* (SCI-EXPANDED) and *Social Sciences Citation Index* (SSCI). WOS contains records of publications from 1900 to the present. The topic terms ‘corpus (corpora)’ and ‘translation’ are used in this step; then 734 records, which range from 2001 to 2021, are generated.

Figure 1 shows the landscape view which is generated on the basis of the publications from 2001–2021. Areas in orange are generated earlier than those in yellow. In what follows, we will focus on the large clusters.



Figure 1. A landscape view of the co-citation network in CTS

As Figure 1 shows, the largest cluster is labeled ‘mediation effect’, and the most actively cited scholar in the cluster is Kruger (2012). The second largest cluster is labeled ‘lexical diversity’, and the academic work which is most frequently cited in the

⁷ <https://citespace.podia.com/>

cluster is Kajzer-Wietrzny (2020). The third largest cluster is labeled ‘Spanish translation’, and the most cited author is Bisiada (2018).

The data show the first three clusters converge to one primary focus in CTS, namely that translated language is a type of constrained language influenced by a mediated effect rather than a bilingual process alone. Based on the colors, the newly activated areas are mediation effect, lexical diversity (or constrained language), theoretical functions, Spanish translation (or editor’s intervention), legal translation, English-German language contact, and medical translation. Among them, Spanish translation (cluster 2), legal translation (cluster 3), medical translation (cluster 7), and English-German language contact (cluster 14) suggest that CTS is cross-fertilizing with neighboring disciplines.

4.1. New emerging research specialties

Based on the visualization and the articles published in the linguistic and translation journals collected from WOS, the new developments that most CTS demonstrate can be briefly interpreted as the main issues which include mediation effect, contact linguistics, and cognitive translation studies.

4.1.1. Mediation effect/constrained language

For a long time, the notion of ‘translation universals’ has been the primary concern in CTS. In line with this notion, studies concerning the mediation effect, constrained language, and editors’ intervention have been recently explored further and developed into the main issues of CTS. Currently, it has been proven that the translated language is primarily the result of a mediation process that is shared among different kinds of mediated language, such as editor’s intervention rather than bilingual processing (Kruger 2012; Kruger and Van Rooy 2016).

Much of what is considered to be universal features of translation are, in fact, features constrained psycho-linguistically and socio-cognitively. This issue is exemplified in Kruger and Van Rooy (2016), who adopt a multi-dimensional approach to analyze a translation corpus and a parallel set of texts from the *International Corpus*

of English (ICE-East Africa)⁸ to determine whether translated and non-native indigenized varieties of English resemble each other. The results demonstrate a shared sets of features between translated and non-native indigenized varieties of English. The study shows that such similarities are the consequence of similar constraints emanating from the cognitive and social environment in which these texts are produced (Kruger and Van Rooy 2016: 27). As a result, translated language is considered as some particular constrained communication. Kaijzer-Wietrzny (2020) further investigates cohesion in the spoken and written registers of constrained language varieties to highlight the similarities and differences in the cohesion patterns of mediated (i.e., interpreted and translated) and non-native texts with respect to original texts produced by native speakers. The study concludes that non-native and mediated texts diverge from native production using cohesive devices in different ways.

4.1.2. Contact linguistics and variational linguistics

Theories related to contact linguistics and variational linguistics have been instrumental in uncovering the motivations behind translated features in CTS. One key finding is that translated language can be seen as a contact variety influenced by language contact. For example, De Sutter and Kruger (2018) have examined the degree of lexico-grammatical explicitness in translated language and compared it to non-contact varieties to determine the factors governing *that*-omission in different types of contact varieties. This approach helps re-evaluate the explanations for the increased explicitness of translated language in the context of language contact. Additionally, CTS can help explore the influence of translation on changes in target languages. In CTS, Pang and Wang (2020) have proposed studying the diachronic correspondence between translated and non-translated texts to uncover the effect of translation on language change. Contact linguistics is a well-established field that employs an electronic methodology and draws on various approaches to predict typical interference from the socio-linguistic and structural description of bilingual communities (Weinreich 1953: 86). The field of contact linguistics covers all linguistic phenomena, such as simplification and other kinds of restructuring, that characterize the outcomes of contact (Winford 2003: 10). Nevertheless, in contact linguistics, there is a paucity of attention to typical properties of

⁸ <https://www.ice-corpora.uzh.ch/en.html>

translations as opposed to non-translations, and such alleged universal properties of translation features are seldom considered.

In recent years, several studies have demonstrated the influence of language contact on translated language and non-translated language. One of them is Kranich *et al.* (2011), who propose ‘language contact through translation’ (LCTT) as a general technical term and present a stepwise approach to studying it by means of corpus methods. More recently, Malamatidou (2017) has further contributed to providing theoretical and empirical perspectives on complex mechanisms that govern the relationship between translation and language change. Likewise, Malamatidou (2018) has also innovatively developed the traditional language contact theories and attempted to integrate the two strands of translation and contact linguistics.

4.1.3. Cognitive translation studies

The combination of Translation and Cognitive Research (TCR) and CTS has yielded some robust findings, with rigorous experimental designs and sophisticated procedures, and has led to several plausible theories in recent years (Alves and Gonçalves 2007; Halverson 2003, 2017). In terms of cognitive explanations, there are two main strands in CTS: the relevance-theoretical account postulated by Alves and Gonçalves (2007) and the cognitive grammatical account proposed in Halverson (2003). Halverson (2003: 50) concludes that many patterns that are proposed to be unique to translation are most likely to be natural effects of bilingual language production, rather than universal characteristics in the translation process.

Similarly, recent studies have harnessed cognitive theories to explain the notion of translation universals, such as the ‘gravitational pull hypothesis’, ‘magnetism’, ‘connectivity’, and ‘general chunking hypothesis’ (Halverson 2017). The gravitational pull hypothesis states that translation characteristics, such as under-representation, can be explained by the structure of semantic networks and prototypes, i.e., the distance between the activated concepts in the semantic network of the bilingual or multilingual translators (De Sutter *et al.* 2017: 2). In line with the study of cognitive translation, Szymor (2018) concludes that the cognitive process underlying the human linguistic system may explain the differences between translated and non-translated texts by comparing the use of core modals in Polish legal texts (originally written in Polish) to

legal texts translated into Polish from English. To be precise, Szymor (2018) claims that it is chunking that has influenced the translators' choice of perfective forms and the authors' choice of imperfective forms.

In translated Chinese, Hu (2019) also demonstrates that the *Ba* structure is more frequently used than in non-translated texts due to the gravitation pull or the third pull in the cognitive system. The study is in line with Ke (2003) and shows that the *Ba* structure is more frequently attested in literary than that in non-literary texts.

4.2. *Emergence of new assumptions and perspectives*

4.2.1. Re-evaluation and re-analysis of the classic topics in CTS

The most classic topic in CTS is syntactic alteration between *that* and zero in English complement clauses (Olohan and Baker 2000). Kruger (2018b) provides a corpus-based multifactorial analysis of the alteration between the retention or omission of *that* in four register-controlled corpora to disentangle the explanations that have been proposed for the increased explicitness of translated English compared to non-translated English texts. Indicators are designed and selected to measure cognitive complexity, pragmatic risk-aversion, and source-language transfer assumptions. Her findings provide strong evidence against the transfer hypotheses and advocate the risk-aversion hypothesis, although the cognitive complexity hypothesis cannot be ruled out.

Similarly, De Sutter and Kruger (2018) analyze English translations from Dutch in Dutch-English parallel corpora and that from Afrikaans, and two self-compiled corpora of written L2 English as the basis to examine the retention and omission of *that*. What makes their study highly interesting is their adoption of the multi-factorial statistical analysis, which incorporates the analysis of the register (which is seen to measure risk-avoidance in respect of formality), the source language structure (to test cross-linguistic influence), and the distance between matrix verb of the main clause and the onset of the complement clause (to measure cognitive effort). The findings show that the choices made in contact varieties are not different from those made in the central variety. The case studies mentioned above represent a methodological advance in CTS and also demonstrate how CTS can benefit methodologically from neighboring disciplines.

4.2.2. Verification of the semantic stability hypothesis in translated texts

For a long time, CTS has been mainly focused on shallow structures in linguistics since such structures are notably marked and easily located, and the research from semantic and pragmatic levels are limited because of the difficulty to operationalize indicators empirically.

As a breakthrough on a semantic level, Vandevoorde (2016) investigates the features of translated texts mainly through the differences between the semantic elements of the verb *start* in translated and non-translated texts by adopting the ‘semantic mirror’ method. The author argues that the semantic meaning in translated text tends to be flattened when compared to non-translated texts. Similarly, De Baets *et al.* (2018) also investigate the semantic features of translated texts through the ‘behavioral profile’. Their study aims to verify the semantic stability hypothesis, namely, whether the semantic structure of an element in translated texts is equal to that in non-translated texts and concludes that the differences between the semantic features in translated texts are lower than those in non-translated texts.

4.2.3. Research on the influence of translation on target language

The study of the relationship between translation and the change of the target language is a key issue in the field of CTS in China and the European countries. In a diachronic study (House 2011), the team led by Juliane House initiated a series of groundwork based on corpora through a ‘covert translation’ project at Hamburg University. It is concluded that text types and social-historical contexts are the main causes for the language changes. Similarly, Kranich *et al.* (2011) shed light on how to combine historical linguistics and CTS.

Pang and Wang (2020) study the translation and change of modern vernacular Chinese based on a new type of corpus, the *Chinese Diachronic Composite Corpus*, which is still under compilation and incorporates a parallel corpus and a comparable corpus in three sampling periods in the twentieth century and a reference corpus as a starting point in the timeframe. They examine whether explicitness in English–Chinese translations has exerted an impact on the target language, focusing on adversative conjunctions as a measure of explicitness. It is proven that the alleged translation universal of explicitation is difficult to substantiate because implicitation and

explicitation both clearly occur in the data in the three sampling periods. This also offers some support for the argument that translated language is constrained by the social context in which the translation takes place. Besides, the combination between translation features and language contact has gained more attention recently. The effect of translation features, such as explicitation/explicitness, shining-through effects, normalization, and under-representation exerted on the change of target languages are examined in the study, which also brings a new perspective to CTS.

4.3. Compilation of multilingual composite corpus and use of statistical techniques

In terms of corpus compilation, while a parallel corpus prevails in cross-lingual variation, a comparable analysis is frequently used to describe translated text features as opposed to non-translated texts. Nevertheless, due to the integration of contrastive linguistics and translation studies, CTS show a trend of the combination of a parallel corpus and comparable corpus. A type of composite corpus, combining both a parallel corpus and a comparable corpus, provides a new insight in cross-linguistic and translation studies —both synchronically and diachronically— by offering a new rationale and methodology to CTS.

In addition, CTS can also be applied to translator and interpreter training, professional practice, translation quality assessment, and machine translation. For instance, the research team at the University of Leuven, led by Sylviane Granger (Granger *et al.* 2018), has made continuous progress in the improvement of bilingual dictionaries. They have also attempted to integrate teaching research and use of corpora to examine the diversity of translators.

CTS has also benefited from advanced research tools, such as *Python* and *R* software. Annotation, statistics, and visualization derived from these tools have provided scholars with more in-depth and insightful findings that could not be noticed and operationalized in early studies. In terms of corpus annotation, the compilation of small corpora with annotation sophisticatedly designed for special uses is also in focus nowadays.

In recent years, corpus-driven and corpus-based studies have paralleled and advanced together. The statistical techniques presented to date may be used in the exploration of the internal and external motivations for the use of the translated features

through text-mining. Cluster analysis, regression analysis, conditional inference tree, and other statistical techniques have been increasingly used.

4.4. Cross-fertilization with other neighboring disciplines

In recent years, CTS has been growing and blending with other disciplines. In addition to contact linguistics and cognitive translation studies (see section 4.1 above), other disciplines, such as discourse analysis and conceptual history, have also contributed substantially to the recent theoretical successes in CTS. Beyond this, the texts examined for CTS are also broadened to medical translation, news translation, legal translation, literary translation, and religious translation.

With regard to discourse analysis, Munday (2012) conducts several case studies on the re-instantiation of appraisal meanings in different genres, based on a corpus-based approach. Similarly, Hu and Meng (2017) coined the term ‘Critical Translation Studies’ and argue that it grows out of the marriage between descriptive translation and critical discourse analysis. Their research aims at investigating ideological factors that operate behind the choice of texts to be translated, the use of translation strategies and methods, features of translated texts, reception of translated texts, and the impact of translation on ideology. In the same vein, Pan and Li (2021) examine the retranslation of political texts —specifically work reports by the Communist Party of China— as a special genre in its own right and contributes to the exploration of the relationship between translation and ideology. By focusing on the retranslation of a recurring set of Chinese political concepts, culture-specific items, and preferred usages into English from the early 1990s to the late 2010s, Pan and Li (2021) showcase how and why the retranslations have been carried out as motivated by the evolving ideologies of the authors.

In the examination of conceptual history, Jones (2020) explores how the shifts in attitudes towards the proper aims and methods of history writing might have shaped the interpretation and translation into English of *Thucydides’ History of the Peloponnesian War*, a work originally written in classical Greek in the fifth century BCE. It initiated the groundwork of the concept studies in the field of translation through the lens of corpora.

It is also worth pointing out that corpus-based methodologies have proven particularly fruitful in the investigation of legal translation. For instance, results from corpus-based analyses can support the description of terminological and phraseological features of legal genres, as well as the acceptability required to make translation decisions and elaborate lexicographical resources in line with legal and institutional translators' needs (Ramos Prieto 2020). Similarly, it is also worth noting that, in recent years, CTS is expanding to Second Language Acquisition and the study of English as a Lingua Franca because both areas focus on bilingual processing.

5. POSSIBLE TRENDS AND FUTURE DIRECTIONS IN CTS

Over the past three decades, CTS has been gradually tearing down the walls between different linguistic disciplines (De Sutter and Kruger 2018) and cross-fertilizing with neighboring disciplines. This has led to equal attention being given to both synchronic and diachronic studies. As De Sutter and Lefer (2019: 2) assert, what is truly essential in CTS is the exploration of corpus-linguistic methods of scrutinizing translational products in order to find the “principles that govern translational behaviour and the constraints under which it operates” (Baker 1993: 235). Motivated by the investigation of the third code, CTS aims to move beyond the traditional examination of indicators, and more theoretical frameworks and disciplines need to be involved. In brief, the possible trends can be framed as shown below.

5.1. *Research aims: From empirical research to theoretical constructs*

As De Sutter and Lefer (2019: 4) suggest, the preoccupation with finding linguistic differences between translated and non-translated texts has left the explanatory framework postulated by Baker—or any other theoretical framework—underdeveloped. Empirical research is conducted from the outset in CTS, but it needs to be guided by theoretical principles to advance when it develops further. Furthermore, since fundamental questions remain largely unanswered (De Sutter and Lefer 2019: 2), theoretical investigations from social, pragmatic, and cognitive mechanisms are expected to provide nutrition for the discipline. In brief, only through theoretical guidelines can we fully understand the phenomenon and unveil the truth of translated features.

Currently, the theoretical construction of CTS can be framed within three directions. These directions encompass cross-linguistic, social, and cognitive explanations of the conditioning forces at play. Exploring neighboring disciplines such as corpus linguistics, linguistic typology, contact linguistics, sociology, Second Language Acquisition, and psycholinguistics can contribute to a more comprehensive understanding of the fundamental principles underlying the nature of translation and the factors that influence it. As CTS advances and enters a new stage of development, new theoretical models are likely to emerge. For instance, Kotze and Havelson (2021) have introduced socio-cognitive constructs in CTS, based on the concepts that linguistic knowledge represents the cognitive organization of an individual's language experience, and normativity involves a feedback loop of conventionalization and legitimization. Their proposed translation model, which connects these two visions, offers an innovative perspective for CTS to explain translation phenomena and understand translators' behavior. This model provides a fresh and insightful framework that enhances the explanatory power of CTS.

5.2. *Research objects: From universality to more constraining factors*

The research on the constraining factors of translation texts needs to involve more variables, such as language contact, cross-lingual comparison, register and translators' expertise, and cognitive factors. The exploration of the universality of translated texts as a third code only serves as a starting point. The goal of this discipline is not restricted to seeking common features in varieties. In contrast, the significance of the third code consists of finding similarities and differences across different varieties to unveil the nature of translation *per se* and all the possible factors shaping constrained communication.

Currently, in line with the traditional topics in CTS, the investigation of the relationship between translation and language change remains an issue of interest in this discipline. The studies of translation texts from the perspective of language contact have yielded a substantial body of research. These studies are in fact proceeded in the 'covert translation' project (House 2011), and in the 'cross-linguistic corpora' (Cro Co) project, led by Silvia Hansen-Schirra, Stella Neumann, and Erich Steiner in Germany (Alves *et al.* 2010). At present, the scope of the studies is gradually expanded, covering linguistic typology, social-linguistics, and cognition.

In terms of constraining factors in language contact situations, Kotze (2020: 346) lists five “overarching and interacting constraint dimensions” that may affect language production: 1) language activation (monolingual–bilingual); 2) modality and register (spoken–written–multimodal); 3) text production (independent/unmediated dependent/mediated); 4) proficiency (proficient–learner); and 5) task expertise (expert–non-expert). Apart from these dimensions, it is expected that more constraining factors will be involved and detected in future studies.

5.3. Research perspectives: From pure linguistics to interdisciplinarity

As noted above, corpus-based translation and interpreting studies are experiencing an upward trend. It is this very interdisciplinary approach that continuously defines CTS and gives it strength to develop and continue to advance. As Halverson (2018) claims, collision of multi-disciplinaries and multi-methods will lead to integration in the next generation of CTS scholar’ research.

For instance, as one important shaping factor, language contact needs more in-depth investigation and verification. The influence of the source language on the target language will help us unveil the nature of translation, as well as the mechanism of code-switching in language change. From the perspective of research methods, the measurement of the similarities and differences between translated and non-translated texts needs more research models and to involve indicators.

In the present time, conjunctions and pronouns are still the main foci of research in CTS, complemented by content words, such as nouns, adjectives, adverbs, cognate words, n-grams, and phraseologies. Besides, new assumptions have also emerged and promoted the study of translation universals. The research perspectives as starting points include the analysis of word chunks, noun phrases, semantic priming, frequency effect, grammatical metaphors, etc., which are not easy to notice. For example, the comparison of the use of verbless sentences in English and Russian and the contrast of linguistic features between directly translated texts and indirectly translated texts also provide insights and highly enrich CTS. In general, the continuous exploration of traditional topics and investigation of new assumptions and perspectives are expected to bring CTS into a new stage.

5.4. Research procedures: From simplification to sophistication

Some important challenges remain in regard to corpus compilation, which is key to CTS. New design of corpora is needed in corpus building, such as multilingual and diachronic composite corpora with more precise semantic and pragmatic annotation. Not only the compilation of large-scale corpora but also of small corpora should be given more attention. In addition, the comparison between written and spoken corpora is also believed to yield new insights into CTS in the future, because it is effective to expand the scope of analysis to the spoken mediated variety.

In terms of statistical analysis, it is expected that frequency-based methods may be replaced by multidimensional, multifactorial, and multivariate statistics. Corpus building and statistical methods serve as one of the driving forces to advance the research in CTS. In the future, it is predicted that statistical methods may be more diversified and sophisticated, such as the use of language modeling statistics by loading and using *R* language packages. Triangulation methods, which combine process and product with empirical data, will be fully exploited and helpful in the future.

Compared to the research conducted in the European countries over the past 20 years, quantitatively speaking, Chinese scholars have provided a large bulk of studies in CTS with many influential pieces of research, such as those by Wang and Qin (2009), Xiao (2010), Hu (2012), and Wang (2012). Although there is always room for Chinese research to improve in theoretical constructs, it has contributed substantially to the compilation of tailor-made corpora, such as the *China English-Chinese Parallel Corpus*⁹ compiled by Beijing Foreign Studies University and the *Political Discourse Corpus*¹⁰ compiled by Shanghai International Studies University, both of which expand the domain of CTS.

6. FINAL REMARKS

This article has provided an overview of the progress that CTS has undergone over the past few decades since its inception in the UK. The development of CTS can be identified into two stages: the establishment of traditional research object in CTS and the expansion of new research issues. Here we have highlighted the current problems

⁹ <http://114.251.154.212/cqp/>

¹⁰ <http://imate.cascorpus.com/>

that persist in the field, including the underestimated complexity of translated features, unresolved conditioning forces, underdeveloped theoretical constructs, and deficient internally coherent corpora with deep annotation. In response to overcoming these problems, the article also presents new developing areas in CTS, such as mediation effects, contact varieties, and cognitive translation studies.

Furthermore, we have identified four trends that have emerged in the field of CTS. First, there has been a shift from empirical research to the development of theoretical constructs in CTS. Second, translation texts are being viewed as contact varieties that are influenced by a range of constraining factors. Third, CTS has adopted an interdisciplinary approach, expanding research perspectives beyond traditional linguistic and translation studies. Finally, there is a growing interest in creating multilingual and diachronic composite corpora and conducting multivariate statistical analyses.

When we look back to what Laviosa (2004: 22) envisaged in CTS at the beginning of this century, it can be said that CTS has not disappointed us in that the potential of corpora in translation studies has been exploited to a large extent. It seems fair to state that, three decades after the publication of Baker's seminal papers, CTS has advanced substantially in the process of converging with neighboring disciplines. At the same time, it has contributed considerably to translation studies, deepening our perception of the nature of translation. We believe what the future holds for CTS is the promotion of interdisciplinary work leading the way towards hybridity and polysemy.

REFERENCES

- Alves, Fabio and José L. Gonçalves. 2007. Modelling translator's competence: Relevance and expertise under scrutiny. In Yves Gambier, Miriam Shlesinger and Rade Gundis Stolze eds. *Doubts and Directions in Translation Studies*. Amsterdam: John Benjamins, 41–45.
- Alves, Fabio, Adriana Pagano, Stella Neumann, Eric Steiner and Silvia Hansen-Schirra. 2010. Translation units and grammatical shifts: Towards an integration of product and process-based translation research. In Gregory M. Shreve and Eric Angelone eds. *Translation and Cognition*. Amsterdam: John Benjamins, 109–142.
- 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 Honour of John Sinclair*. Amsterdam: John Benjamins, 233–252.
- Becher, Viktor. 2010. Abandoning the notion of 'translation-inherent' explicitation: Against a dogma of translation studies. *Across Languages and Cultures* 11/1: 1–28.

- Bisiada, Mario. 2013. *From Hypotaxis to Parataxis: An investigation of English-German Syntactic Convergence in Translation*. Manchester: The University of Manchester Dissertation.
- Bisiada, Mario. 2018. The editor's invisibility: Analyzing editorial intervention in translation. *Target* 30/2: 288–309.
- Chen, Chaomei. 2017. Science mapping: A systematic review of the literature. *Journal of Data and Information Science* 2/2: 1–40.
- De Baets, Pauline, Lore Vandevordee and Gert De Sutter. 2018. Meaning shifts in translation: A corpus-based behavioural profile approach. In Sylviane Granger, Marie-Aude Lefer and Laura Aguiar de Souza Penha-Marion eds., 47–49.
- De Sutter, Gert and Haidee Kruger. 2018. Disentangling the motivations underlying syntactic explicitation in contact varieties: A MuPDAR analysis of *that* vs. zero complementation. In Sylviane Granger, Marie-Aude Lefer and Laura Aguiar de Souza Penha-Marion eds., 55–57.
- De Sutter, Gert and Marie-Aude Lefer. 2019. On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives* 28/1: 1–23.
- De Sutter, Gert, Bert Cappelle, Orphée De Clercq, Looock Rudy and Koen Plevoets. 2017. Towards a corpus-based statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translation. *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 16: 25–39.
- Dimitrova, Englund. 2005. *Expertise and Explicitation in the Translation Process*. Amsterdam: John Benjamins.
- Frawley, William. 1984. Prolegomenon to a theory of translation. In Lawrence Venuti ed. *The Translation Studies Reader*. London: Routledge, 250–263.
- Granger, Sylviane, Marie-Aude Lefer and Laura Aguiar de Souza Penha-Marion eds. 2018. *Book of Abstracts: Using Corpora in Contrastive and Translation Studies*. Louvain-la-Neuve: Centre for English Corpus Linguistics and Université Catholique de Louvain.
- Halverson, Sandra. 2003. The cognitive basis of translation universals. *Target* 15/2: 197–241.
- Halverson, Sandra. 2017. Developing a cognitive semantic model: Magnetism, gravitational pull and questions of data and method. In Gert De Sutter, Marie-Aude Lefer and Isabelle Delaere eds. *Empirical Translation Studies: New Methodological and Theoretical Traditions*. Berlin: Mouton de Gruyter, 9–45.
- Halverson, Sandra. 2018. Cognitive translation studies and the combination of data types and methods. In Sylviane Granger, Marie-Aude Lefer and Laura Aguiar de Souza Penha-Marion eds., 3–5.
- Hermans, Theo. 1999. *Translation in Systems: Descriptive and System-oriented Approaches Explained*. Manchester: St Jerome.
- House, Juliane. 2011. Using translation and parallel text corpora to investigate the influence of global English on textual norms in other languages. In Alet Kruger, Kim Wallmach and Jeremy Munday eds. *Corpus-based Translation Studies: Research and Applications*. London: Continuum, 187–208.
- Hu, Kaibao. 2012. *Introduction to Corpus-based Translation Studies*. Shanghai: Shanghai Jiao Tong University.
- Hu, Xianyao. 2019. Re-examining *ba* structure in translated Chinese. Presentation at the *Fifth Conference of Corpus-based Translation Studies*. November 2019 Chongqing: Chongqing university.

- Hu, Kaibao and Lingzi Meng. 2017. Critical translation studies: New development in translation studies. *Journal of Foreign Languages* 40/6: 57–68.
- Jones, Henry. 2020. Retranslating *Thucydides* as a scientific historian: A corpus-based analysis. *Target* 32/1: 59–82.
- Kajzer-Wietrzny, Marta. 2018. Translationese, interpretese and foreignese: What do they have in common? In Sylviane Granger, Marie-Aude Lefer and Laura Penha-Marion eds. *Book of Abstracts. Using Corpora in Contrastive and Translation Studies Conference*. Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain, 93–94.
- Kajzer-Wietrzny, Marta. 2020. A multivariate approach to lexical diversity in constrained language. *Across Languages and Cultures* 21/2: 169–194.
- Ke, Fei. 2003. Features and dispersion of *ba* structure in Chinese texts and its Chinese-English translation. *Foreign Languages and their Teaching* 177/12: 1–5.
- Ke, Fei. 2005. Implication and explicitation in translation. *Foreign Language Teaching and Research* 37/4: 303–307.
- Kenny, Dorothy. 2001. *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester: St. Jerome.
- Klaudy, Kinga and Krisztina Károly. 2005. Implication in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures* 6/1: 13–28.
- Kotze, Haidee. 2020. Translation, contact linguistics and cognition. In Fabio Alves and Arnt Lykke Jakobsen eds. *The Routledge Handbook of Translation and Cognition*. Abingdon: Routledge, 113–132.
- Kotze, Haidee and Sandra Havelson. 2021. Norms, constraints, risks: A usage-based perspective on sociocognitive constructs in corpus-based translation studies (and beyond). In Sara Castagnoli, Silvia Bernardini, Adriano Ferraresi and Maja Miličević Petrović eds. *Book of Abstracts: Using Corpora in Contrastive and Translation Studies Conference*. Bertinoro: University of Bologna, 189–190.
- Kranich, Svenja, Viktor Becher, Steffen Hoder and Juliane House eds. 2011. *Multilingual Discourse Production*. Amsterdam: John Benjamins.
- Kruger, Haidee. 2012. A corpus-based study of the mediation effect in translated and edited language. *Target* 24/2: 355–388.
- Kruger, Haidee. 2018a. *That* again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures* 20/1: 1–33.
- Kruger, Haidee. 2018b. Expanding the third code: Corpus-based studies of constrained communication and language mediation. In Sylviane Granger, Marie-Aude Lefer and Laura Aguiar de Souza Penha-Marion eds., 9–12.
- Kruger, Haidee and Bertus Van Rooy. 2016. Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide* 37/1: 26–57.
- Laviosa, Sara. 2004. Corpus-based translation studies: Where does it come from? Where is it going? *Language Matters* 35/1: 6–27.
- Malamatidou, Sofia. 2017. Creativity in translation through the lens of language contact: A multilingual corpus of *A Clockwork Orange*. *The Translator* 23/3: 1–18.
- Malamatidou, Sofia. 2018. *Corpus Triangulation: Combining Data and Methods in Corpus-based Translation Studies*. London: Routledge.

- Munday, Jeremy. 2012. New directions in discourse analysis for translation: A study of decision-making in crowdsourced subtitles of Obama's 2012 state of the Union speech. *Language and Intercultural Communication* 12/4: 321–334.
- 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.
- Øverås, Linn. 1998. In search of the third code: An investigation of norms in literary translation. *Meta* 43/4: 557–570.
- Pan, Feng and Tao Li. 2021. The retranslation of Chinese political texts: Ideology, norms, and evolution. *Target* 33/3: 381–409.
- Pang, Shuangzi and Kefei Wang. 2020. Language contact through translation: The influence of explicitness in English–Chinese translation on language change in vernacular Chinese. *Target* 42/3: 420–455.
- Ramos Prieto, Fernando. 2020. Translating legal terminology and phraseology: Between inter-systemic incongruity and multilingual harmonization. *Perspectives* 29/2: 175–183.
- Redelinghuys, Karien and Haidee Kruger. 2015. Using the features of translated language to investigate translation expertise: A corpus-based study. *International Journal of Corpus Linguistics* 20/3: 293–325.
- Shneider, Alexander M. 2009. Four stages of a scientific discipline: Four types of scientists. *Trends in Biochemical Sciences* 34/5: 217–223.
- Szmrecsanyi, Benedikt and Bernd Kortmann. 2009. Vernacular universals and angloversals in typological perspective. In Markku Filppula, Juhani Klemola and Heli Paulasto eds. *Vernacular Universals and Language Contact: Evidence from Varieties of English and Beyond*. New York: Routledge, 33–53.
- Szymor, Nina. 2018. Translation: Universals or cognition? A usage-based perspective. *Target* 30/1: 53–86.
- Vandevoorde, Lore. 2016. *On Semantic Differences: A Multivariate Corpus-based Study of the Semantic Field of Inchoativity in Translated and Non-translated Dutch*. Ghent: Ghent University dissertation.
- Wang, Kefei. 2012. *Exploring Corpus-based Translation Studies*. Shanghai: Shanghai Jiao Tong University Press.
- Wang, Kefei and Hongwu Qin. 2009. A parallel corpus-based study of general features of translated Chinese. *Foreign Language Research* 146: 102–105.
- Weinreich, Uriel. 1953. *Language in Contact: Findings and Problems*. The Hague: Mouton de Gruyter.
- Winford, Donald. 2003. *Contact Linguistics*. Oxford: Blackwell.
- 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.
- Zanettin, Federico. 2012. *Translation-Driven Corpora*. Manchester: St. Jerome.

Corresponding author

Kefei Wang

Beijing Foreign Studies University

National Research Center for Foreign Language Education

North Xisanhuan Avenue

PO Box 89-45

100089 Beijing

China

Email: kfwang@bfsu.edu.cn

received: November 2022

accepted: February 2023