

Empirical Translation Studies

Introducing Corpora in Translation Studies: 1993–2003

The introduction of corpora in Translation Studies was conceived within an empirical paradigm and came to be because of the convergence between the discovery and justification procedures put forward by Gideon Toury (1995/2012) for the study of translation and the data-driven approach developed by Corpus Linguistics for the study of languages. The synergy between Descriptive Translation Studies and Corpus Linguistics acted as a stimulus to the creation of a variety of corpus resources, the development of a descriptive research methodology and the growth of a line of enquiry that was put forward in the 1980s and gathered momentum thanks to the availability of corpora. This body of research is known as the quest for translation universals, which are posited as probabilistic laws of translational behaviour (Toury 1995/2012).

One of the first corpus resources designed for contrastive linguistics and translation studies is the English-Norwegian Parallel Corpus (ENPC). It was compiled at the University of Oslo under the direction of Stig Johansson and served as a model for the bidirectional parallel corpus of English and Portuguese, COMPARA.2 Another corpus design is the monolingual comparable corpus. An example is the Translational English Corpus (TEC).3 It was created at the University of Manchester under the direction of Mona Baker. Another example is the Corpus of Translated Finnish (CTF). It was compiled at the Savonlinna School of Translation Studies by Anna Mauranen's research group. The methodology adopted by CTS involves a helical progression from the elaboration of descriptive, interpretive and explanatory hypotheses to inferences about the non-observable culturally determined norms that govern translators' choices.

Initially, research focused on four universals: simplification, explicitation, the law of growing standardization (largely compatible with normalization) and the law of interference. Simplification is "the process and/or result of making do with less words" (Blum-Kulka and Levenston 1983: 119). Explicitation is "an observed cohesive explicitness from SL to TL texts regardless of the increase traceable to differences between the linguistic and textual systems involved" (Blum-Kulka 1986: 19). The law of growing standardization posits that "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire" (Toury 1995/2012: 304). The law of interference states that "in translation, phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred to the target text" (Toury 1995/ 2012: 310). In sum, during the first decade of its life, CTS built upon, refined, extended, and diversified previous research into the regularities of translational language.

Meanwhile, corpora were making inroads into Applied Translation Studies. In this area of research and practice, corpora were used mainly as translation aids in translator training. Corpora were utilized as repositories of data for retrieving translation equivalents, acquiring content knowledge about specialized subject fields and developing stylistic fluency and terminological accuracy in the target language. Translation pedagogy drew mainly on Data-

Driven Learning (DDL), developed by Tim Johns for the teaching of languages (Johns 1991a, b), and on constructivist principles, which constitute a dominant paradigm in contemporary educational philosophy and “serve as a strong cornerstone for the development of student- and praxis-relevant teaching methods” (Kiraly 2003: 8).

More specifically, the DDL approach adopts the principles of Corpus Linguistics and involves carrying out small-scale projects where students identify problem areas arising from translation practice, suggest hypotheses and then test them with their own tutor who has the role of “director and coordinator of student-initiated research” (Johns 1991a: 3). The approach adopted by the collaborativeconstructivist method combines social constructivism with modern functionalist theories and expertise studies. The design involves collaborative learning and project-based activities. The procedure requires that students engage in an authentic or realistically simulated translation project together with peers (Kiraly 2000, 2003). Summing up, within the empirical paradigm, which can be regarded, in line with Chesterman (1998), as the most important trend that characterized Translation Studies in the 1990s, corpora engendered a number of novel syntheses in the pure and applied branches of the discipline.

Consolidating Corpora in Translation Studies: 2003–2013

The second decade in the life of CTS is marked by two international conferences entirely devoted to corpora and Translation Studies. The first was held in Pretoria in 2003, it was entitled *Corpus-based Translation Studies: Research and Applications* (Kruger et al. 2011). The second was hosted in Shanghai in 2007, *Conference and Workshop on Corpora and Translation Studies*. Of note is also the establishment of a strong partnership between contrastive and translation studies, in keeping with the research programme initiated by Stig Johansson in the 1990s and pursued in several interdisciplinary collected volumes such as Granger et al. (2003).

As regards Descriptive Translation Studies, many new corpora have been created in the last ten years or so, as amply testified by Federico Zanettin’s web page⁴ and his monograph (Zanettin 2012). A novelty is the design of corpora of interpreted speeches, the first being the European Parliament Interpreting Corpus (EPIC),⁵ and the consequent growth of a new body of research named *Corpus-based Interpreting Studies (CIS)* (Setton 2011). Its main goal is to unearth the specificity of interpreting vis-à-vis original oral discourse and written translation in the same target language. This is a line of enquiry that was first proposed and pursued by Shlesinger (1998, 2009). The creation of new corpus resources goes hand in hand with the development of methodology and statistical analyses. These are enriched by contextual data and are becoming more and more sophisticated thanks to the advancement of technology. As for the range of research endeavours, the quest for translation universals goes on (Mauranen and Kujamäki 2004; Russo et al. 2006; Espunya 2007; Bruti and Pavesi 2008; Steiner 2005, 2008; Ulrych and Anselmi 2008; Gaspari and Bernardini 2010; Xiao et al. 2010; Zuffery and Cartoni 2014). In addition to the four universals mentioned earlier, a new one emerged, namely the Unique Items Hypothesis (UIH). It states that target-language-specific elements, which do not have equivalents in the source language, tend to be under-represented in translated texts, since “they do not readily suggest themselves as translation equivalents” (Tirkkonen-Condit 2004: 177–178).

Despite growing interest in this ambit of research, a number of scholars expressed serious criticism about the tenability of the very concept of translation universal. House (2008), for example, claimed that the investigation of universals is futile since there are no and there can be no translation-inherent universals. The reasons for denying the existence of linguistic features of translation per se are as follows: (a) since translation is an act that operates on language, the universals of language also apply to translation; (b) translation is inherently language-pair specific, hence corpus-based multi-pair comparisons remain agglomerations of different pairs; (c) the suggested candidates for the status of translation universal for one particular translation direction need not necessarily be candidates for universality in the opposite direction; (d) translation universals have been found to be genre-sensitive, for instance, while there is a tendency towards explicitation in German translation of popular scientific texts, this is not the case to the same degree for economic texts; (e) translations may be influenced by the status of the language of the source text genre, which in turn may influence the nature of the translation text genre and also the nature of comparable texts in the same genre (House 2008: 11). Becher (2011) shares House's critical stance. The departure point for his corpus study of English-German and German-English translations of business texts was not the assumption that explicitation is a translation-inherent universal process. Instead, Becher predicted that every instance of explicitation (and implicitation) can be accounted for by lexicogrammatical and/or pragmatic factors. His findings confirmed this hypothesis.

Malmkjær's (2008) contribution to the ongoing debate on the posited existence of translation universals is both critical and constructive. She suggests that universals such as simplification, explicitation and normalization would be better accounted for by the norm concept and explained on socio-cultural grounds. Instead, the Unique Items Hypothesis (UIH) is a good candidate for universal status because it can be explained on cognitive grounds. Indeed, the UIH, which has been confirmed by studies carried out with unrelated languages (Swedish and Danish on the one hand, and Finnish on the other), is a phenomenon that is not triggered by the source text, but seems to arise during the translation process, from the under-representation in a translator's mental lexicon of unique features of the target language. Malmkjær argues that, if the concept of the translation universal is to retain any theoretical credibility, it would have to be reserved to phenomena such as the UIH, "for which it makes sense to produce a cognitively based explanation" (Malmkjær 2008: 57). In addition to the quest for translation universals, other research projects were pursued during the 2003–2013 decade. They concern the style of literary translators, the role of ideology in determining translation choices and the study of Anglicisms.

In Applied Translation Studies corpora continued to be used to retrieve and examine lexical, terminological, phraseological, syntactic and stylistic equivalents. They also began to be utilized in Translation Quality Assessment (Bowker 2003a, b). Corpora have been increasingly incorporated in the curricular design of postgraduate translator training programmes to satisfy the exigencies of today's globalized and technologized language industry (Koby and Baer 2003; Zanettin et al. 2003; Kelly 2005; Ulrych 2005; Olohan 2007; Rodrigo 2008; Beeby et al. 2009). In sum, we can affirm that a coherent interdisciplinary theory combined with the professional and institutional recognition of corpora as valuable linguistic resources and translation aids has given rise to an effective partnership that is playing a crucial role in engendering a culture of research in education.

The question to be addressed at this point is the extent to which corpora have provided “an opportunity to reengage the theoretical and pragmatic branches of Translation Studies, branches which over and over again tend to disassociate, developing slippage and even gulfs”, as was envisaged by Tymoczko (1998). The relationship between corpus-based descriptive and applied studies has been open and reciprocal to a degree. As we saw earlier, this line of enquiry was conceived as a descriptive research endeavour. Its findings were then projected into Applied Translation Studies where the Unique Items Hypothesis was tested and confirmed experimentally in the undergraduate translation classroom to raise awareness among students of what translation entails (Kujamäki 2004). Also, simplification and explicitation were tested as possible indicators of translation quality with a view to improving teaching methods and assessment criteria at postgraduate level. Simplification was found to correlate with lower-scoring translations and explicitation was found to correlate with higher-scoring ones (Scarpa 2006).

These studies, which engage in classroom-based investigations inspired by the insights provided by the pure branch of the discipline, represent the beginning of a new trend in Translation Studies. Moreover, it is a promising orientation not only because it aims to replicate descriptive investigations and render translation teaching more effective and evaluation more rigorous, but also because it empowers students to gain a deep and critical understanding of the process, product, and function of translation. Thanks to this knowledge, translator trainees will be capable of adhering to or innovating culturally determined norms in an informed, conscious, and responsible way. As Pekka Kujamäki contends, theories, models, concepts, and experimentation with students should have an essential role in translation pedagogy “not only in research seminars but also and above all in the translation class: they open a way to novices’ better understanding of their future status as experts of human translation” (Kujamäki 2004: 199). In line with this envisioned direction for CTS, the holistic approach to translating culture elaborated by Tymoczko (2007) should be adopted as a theoretical framework within which corpora can reengage the pure and applied branches of the discipline for the benefit of both.

Looking to the Future: Corpora and Holistic Cultural Translation

The holistic approach to translating cultural difference presupposes that translation be conceived as an open, cluster concept with blurred edges. This notion of translation, which Tymoczko calls the ‘cluster concept translation’ (or ‘translation with an asterisk’), is defined in terms of resemblances between translation and three forms of cultural interface, i.e., representation, transmission, and transculturation. we shall now define each of these large superordinate categories that partially encompass, impinge on, and illuminate translation.

As a form of representation, translation offers an image or likeness of another thing, it stands in place of another entity and has authority to substitute for or act in place of that entity. Almost all translations are forms of representation, with a few exceptions such as pseudo translations (or fictitious translations) (Toury 1995/2012: 47–59). As a form of transmission, translation involves different types of transfer from one language and culture to another. Translations typically relay the content, language, function or form of the source text. The variability of methods adopted by translations that privilege transfer is very wide. It ranges from close textual

fidelity to various degrees of manipulation of the linguistic features of the original. Many factors influence the vast array of transmission procedures adopted in translation practices, e.g. linguistic asymmetries, translation technologies, literacy practices, economic conditions, cultural sufficiency or enclosure, receptiveness to difference, aesthetic norms, taboos about certain types of content, asymmetries in power and cultural prestige as well as ideology (Tymoczko 2007: 119).

As a form of transculturation, translation is the transmission and uptake of borrowed cultural forms in the receptor environment and the consequent creation of new cultural phenomena. Transculturation generally includes such elements as verbal materials, religious beliefs and practices, social and political organization, artistic forms as well as aspects of material culture including technology and tools, agricultural and industrial practices, clothing, food, housing, transport and media (Tymoczko 2007: 120). In textual domains, transculturation often entails transposing elements of a literary system, e.g., poetics, genres, tale types. It also involves the uptake of the elements expressed in or carried by language such as discourses and world views (2007: 121). Lexical borrowing from a donor to a receptor language is a form of transculturation that occurs through language contact. Philip Durkin's historical linguistic study of loanwords in English (2014) shows how the rich variety of the English lexis reflects the vast number of words it has borrowed from languages as varied as Latin, Greek, Scandinavian, Celtic, French, Italian, Spanish, Russian Hebrew, Māori, Malay, Chinese, Hindi, Japanese and Yiddish. Also, the study of Anglicisms, a linguistic and socio-cultural phenomenon that has been investigated extensively through corpora, has attracted the interest of a growing number of scholars in recent years, particularly in Europe, where there is a need of reconciling the role of English as a European lingua franca with the EU commitment to cultural and linguistic diversity (Anderman and Rogers 2005; Furiassi et al. 2012; House 2012).

An insightful corpus-based study of translation as a form of transculturation is Ji's (2013) historical linguistic investigation of the introduction, assimilation and appropriation in early modern Chinese language and culture of the western key term nation in its modern sense of 'nation state'. The study covers five decades: 1840-1850, 1860-1870, 1880-1890, 1900-1910 and 1910-1920. The results show that during the 1840-1850 decade, the term nation was translated from English, French, Dutch, German and Italian mostly by words denoting the ethnic composition of a nation (i.e. zhonglei, zulei, yizu, zuzhong), while the word guomin, which referred to the people representing a sovereign state, was significantly less frequent. But in the subsequent decades this pattern changed. The most frequent equivalents of nation became guo and bang, which denoted a geopolitical entity. Less frequent were the words used to refer to the people of a given country (e.g. min and guomin).

The cluster concept translation rests on the assumption that language and culture are closely intertwined and "culture is the domain where human differences are most manifest" (Tymoczko 2007: 221). When communicating across cultural differences, as Tymoczko argues, it is not sufficient to approach the representation of culture in a linear, piecemeal fashion and resolve the problems incorporated in surface elements of the text one by one, sentence by sentence until the translation is complete (Tymoczko 2007: 233). What is needed instead is a holistic approach. As Tymoczko explains, "a holistic approach to translating culture will begin with the largest elements of cultural difference that separate the source culture and the target

culture as a framework for coordinating the particular decisions about culture that occur as the text is actually transposed into the target language” (Tymoczko 2007: 235).

In order to help translators accomplish such a complex task, Tymoczko offers a partial repertory of cultural elements that might be taken into account as a guide for interpreting the source text and for determining the overall representation of culture in the target text. The inventory comprises: • Signature concepts of a culture • Key words • Conceptual metaphors • Discourses • Cultural practices • Cultural paradigms • Overcodings • Symbols.

We will now define each of these large cultural elements in turn and illustrate them with examples from various languages and genres. As we shall see, many of these examples are offered by corpus research. Signature concepts express key values in the social and economic organization of a culture. The words denoting them are highly connoted and rich in cultural associations. In early medieval Irish texts, for instance, words belonging to the semantic field of heroism, such as honour, shame, taboo, fall under the category of signature concepts (Tymoczko 2007). The signature concepts of contemporary American society can be equated to the values that American citizens cherish and are encouraged to promote. These are “hard work and honesty, courage and fair play, tolerance and curiosity, loyalty and patriotism”, as we read in the letter that the President of the United States of America sends to every new American citizen. On the other side of the Atlantic, the liberal values held by British people today are openness, tolerance, compassion, and strength.

Key words are words that may point either to the signature concepts of a culture or to the thematic cultural elements chosen by a writer or speaker to structure a given text or a corpus of texts. For example, the strongest key words analyzed by Norman Fairclough (2000) in the corpus of New Labour texts (which contains a variety of texts produced under the New Labour Government led by the British Prime Minister Tony Blair from 1994 to 1999) are: New Labour, new deal, new Britain, business and partnership, welfare reform (Fairclough 2000:17–20).

Drawing on the work of cognitive linguists such as Lakoff and Johnson (1980), conceptual metaphors shape the way we understand and experience reality, i.e. they structure (at least in part) what we do and how we understand what we are doing (Lakoff and Johnson 1980: 5). An example of variation in conceptual metaphors across languages is offered by Ding et al. (2010). Their corpus-based analysis of the metaphorical representation of the topic FEAR and its Chinese equivalent KONGJU, reveals that Chinese does not have the English conceptual metaphors FEAR IS A SUPERNATURAL BEING/A DISEASE/A SHARP OBJECT/A POISON/A LEGACY/A MACHINE. Moreover, the shared metaphor FEAR IS AN OPPONENT tends to be used in English to conceptualize the state of falling victim to fear, whereas in Chinese it is usually used to conceptualize an attempt to control it.

Ideological discourses are representations and visions of the social world expressed in speech and writing and as such they motivate action and cultural practice. They are the object of study of Critical Discourse Analysis, an area of research which has been investigated extensively through corpora. An example of such analysis is offered by Fairclough’s research into the political discourse of the ‘Third Way’ in Tony Blair’s speeches from 1998 to early 1999. The

Third Way signifies a programme that was defined by centre and centre-left British governments as being neither old left nor 1980s right. It was built upon the notion of “the new global economy”, that was accepted “as an inevitable and unquestionable fact of life” upon which politics and governments were to be premised (Fairclough 2000: 15,150). In the European Union the discourse of ‘unity in diversity’, which first came into use in 2000 “signifies how Europeans have come together, in the form of the EU, to work for peace and prosperity, while at the same time being enriched by the continent’s many different cultures, traditions and languages.”

Cultural practices such as naming practices, forms of address and titles, the naming of kinship relationships play an important role in constructing personal and social identities and achieving social cohesion. They too may vary across languages. In English, for example, the word grandfather means ‘father of one’s father or mother’ and the word grandmother means ‘mother of one’s father or mother’. The Italian equivalents are: nonno and nonna respectively. But in Thai the word ป (po) means ‘father of one’s father’, the word ตา (ta) means ‘father of one’s mother’, the word ย (ya) means ‘mother of one’s father’, and the word ยาย (yay) means ‘mother of one’s mother’. Similarly, in Swedish farfar = father’s father, morfar = mother’s father, mormor = mother’s mother and farmor = father’s mother. In Chinese there are five equivalents of the English word uncle, i.e. shushu, bobo, jiujiu, guzhang, andyizhang, each referring to a specific family relationship.

Cultural paradigms pertain to humour, argumentation, logical sequencing in a text or the use of tropes. They tend to vary from language to language and within the same language over time. For example, a corpus-based study carried out by Niu and Hong (2010) on rhetorical repetition in English and Chinese print ads published in two leading newspapers in Singapore shows different patterns. The four most frequent repetition types in English are alliteration, rhyme, assonance, anaphora, while in Chinese they are assonance, anaphora, alliteration, rhyme. So, the results show that English uses more alliteration and rhyme than Chinese, and Chinese uses more assonance and anaphora than English in this particular genre.

Overcodings are “linguistic patterns that are superimposed on the ordinary ranks of language to indicate a higher-order set of distinctions in language practices” (Tymoczko 2007: 243). They signal specific literary genres (e.g., poetry or narrative) and modes of communication (e.g., spoken or written). They also comprise rhetorical devices such as intertextuality, quotation, and allusion. For example, the literary style of Latino writers in the United States is characterized by a constant code-switching from English to Spanish. As Díaz Pérez (2012:171–172) observes, “[b]y introducing Spanish words, phrases or syntactic constructions into their English texts, they try to evoke the feeling of living on a frontera, of inhabiting two worlds which can be conflicting and complementary at the same time.”

Within the category of overcodings we also find forms of textual structuring pertaining to aspects of register, dialects and languages for special purposes. An excellent example of corpus-based research that throws light into the relationship between overcodings and cultural context is Meng Ji’s investigation of the lexico-grammatical features that characterize scientific language in early modern Chinese. This specialized register was developed by translating

scientific texts from Western languages, most notably English, French and Dutch from the mid-nineteenth century to the turn of the twentieth century. This was a time characterized by the expansion of capitalism and imperialism in Asia. Two types of overcodings were unveiled by Ji's study: (a) disyllabic word structure, i.e., words created by combining two existing characters, and (b) functional particles.

Functional particles were created in Chinese to relay the meanings and functions expressed by the prefixes and suffixes of Latin and Greek origin that characterized Western scientific writing. Ji's study reveals two groups of functional particles, i.e. grammatically modified and semantic-cognitive functional particles. An example of the former is *de*, which identifies an adjective and was retrieved from ancient Chinese literary fiction. An example of the latter is *zhe*, an abstract term for things, agents or concepts, which was retrieved from ancient philosophical and historical texts as well as biographical essays. As Ji observes, while the original affixes "reflect the systematicity and continuity of the development of modern scientific language based on ancient Latin and Greek cultures and thoughts" (Ji 2012: 255), the development of equivalent functional particles in early Chinese scientific language "involved a thorough and painstaking re-examination of the target language body, searching for expressions of metaphorical references parallel to their Western counterparts" (Ji 2012: 255).

Finally, symbols are related to the identity of an individual, family, class, nation or deity (Tymoczko 2007: 145, fn. 28). Indeed, flower symbolism varies from language to language. Lilacs stand for light and early summer in Sweden but in Italy they represent envy. In some English villages a lilac branch may signify a broken engagement (Anderman 2007: 3). Folklore provides many other symbols and icons. In Indian mythology the word *naga* describes any kind of semi-divine serpent associated with water and fluid energy. Nagas are ambivalent deities, they are believed to bestow wealth and assure abundant crops but revoke these blessings if offended. An example of a symbol that crosses cultural boundaries is the poppy flower. During the years that followed the Great War it was adopted as the symbol of remembrance, especially in the Commonwealth countries. The idea was inspired by a poem written by the Canadian military doctor and artillery commander Major John McCrae in memory of his friend, a young Canadian artillery officer, Lieutenant Alexis Helmer, who was killed on 2nd May, 1915 when an exploding German artillery shell landed near him during the early days of the Second Battle of Ypres (Corni and Fimiani 2014: 307).

As Tymoczko maintains, considering all the above cultural elements helps translators compare their own culture with the source culture as it is reflected in texts. To make these cross-cultural comparisons translators need to develop self-reflexivity. It is through self-reflexivity that they will be able to identify those elements of cultural difference that need to be mediated. As a result, "a holistic approach to cultural translation rather than a selective focus on a limited range of cultural elements enables greater cultural interchange and more effective cultural assertion in translation, allowing more newness to enter the world" (Tymoczko 2007: 233). Corpora can play an important role in fostering holistic cultural translation since they can fruitfully be used "to illuminate both similarity and difference and to investigate in a manageable form the particulars of language-specific phenomena of many different languages and cultures" (Tymoczko 1998: 657).

Towards a Corpus-Based Holistic Pedagogy

How can corpora be used to unearth cross-cultural differences and similarities in research as in practice? An effective way of achieving this goal is to work towards a multilingual pedagogy that espouses the tenets of holistic cultural translation and incorporates corpora not only as tools for acquiring technical skills, but also as resources for developing translingual and transcultural competences that enable translators to act as self-reflexive, responsible meaning makers in our increasingly globalized, multicultural world. In such envisioned pedagogy comparable and parallel corpora should be explored through discovery and justification procedures (Toury 1995/2012) so as to infer culturally determined norms and particularities on the basis of empirical evidence.

What follows is an example of how this teaching method was adopted to investigate a form of transculturation, i.e. the relationship between the loanword *business* in Italian and its English etymon. Loanwords are problematic in translation as their lexico-grammatical profiles tend to be different across donor and receptor languages. An Anglicism may, for example, convey only a subset of the senses expressed by the English etymon. Also, an Anglicism may acquire different connotations in the receiving language. As Pulcini (2002: 162) explains, lexical borrowing is a complex phenomenon “because it involves referential, connotative, contextual and sociocultural components of meaning”. Consequently, the senses conveyed by words in the donor language may be “kept, altered, restricted or expanded” in the receiving language (Pulcini 2008: 196). Normally, if an English word is borrowed in order to fill a semantic gap in Italian, the referential meaning remains the same, as is the case with the terms *agribusiness* or *bed and breakfast*. But in many other instances, changes tend to occur in the form of restriction or expansion.

An example of restriction is offered by the term *benchmark*. In Italian it refers to a financial market index that enables investors to assess the upward or downward trend of an investment fund (Pulcini 2008: 197). In English *benchmark* is used as a verb and a noun. As a verb it means “to provide a standard that something can be judged by”. As a noun it means “an amount, level, standard etc. that you can use for judging how good or bad other things are” (Macmillan English Dictionary for Advanced Learners 2007). An example of expansion of meaning and change of word class is provided by the borrowed term *backstage*. As an English adverb it means “in the area behind the stage in a theatre, including the rooms where the actors get dressed”. As an adjective it means “relating to the area behind the stage in a theatre, including the rooms where the actor get dressed: a backstage pass (= a special ticket that allows you to go backstage)” (Macmillan English Dictionary for Advanced Learners 2007). In Italian the noun *backstage* refers to the area behind the stage in a theatre and to a documentary that illustrates the technical problems, atmosphere, emotions, and gossip involved in the preparation of a film, event or theatre performance (Pulcini 2008: 198).

These cross-linguistic asymmetries largely arise from the fact that “meanings are established in individual languages by contrasts of similar items in semantic fields” (Görlach 2003: 93). Hence, translator trainees working out of English often find it difficult to decide when to use Anglicisms appropriately. Since translated texts can serve as semantic mirrors reflecting meaning across languages (Johansson 2003: 136), they are an invaluable resource for investigating loanwords across donor and receptor languages. Before moving on to report on a small-scale research project conducted in the postgraduate translation classroom to raise

awareness of translation norms and the phenomenon of lexical borrowing, we will outline the procedural steps we adopted in keeping with the methodology elaborated by Toury (1995/ 2012) for discovering regularities in translational behaviour.

Toury's Discovery and Justification Procedures

The research methodology proposed by Toury (1995/2012:31–34) for descriptive translation studies is articulated in three phases. The first phase starts with the selection of individual translations or a corpus of translations within the target culture. Toury's perspective is target-oriented: translations are texts that belong to the target culture, they are texts in their own right, not just mere representations of their source texts. The analysis carried out in the first phase involves the initial assessment of the acceptability of the individual translations or corpus of translated texts without reference to the source texts. Acceptability is the extent to which a translated text adheres to the linguistic and cultural norms prevailing in the target language for a particular text genre. The opposite concept is that of adequacy. An adequate translation is one which leans towards the norms of the source language and culture and contains traces of the textual features of the source text.

The second phase starts with the identification of the source texts and proceeds to comparing the target texts and their sources in parallel, that is sentence by sentence, paragraph by paragraph. The aim is to determine target-source relationships, translation problems, translation solutions and shifts. According to Toury they can be of two kinds: obligatory, which are caused by systemic differences between the source and target languages, and non-obligatory, which are motivated by literary, stylistic or cultural considerations.

In the third phase of the analysis, the relationships established between the target texts and their sources become the basis of first-level generalizations about the initial norm underlying the concrete way in which equivalence is realized. The initial norm governs the basic choice which can be made between adequacy (which involves adhering to source norms) and acceptability (which involves subscribing to norms originating in the target culture). Adequacy and acceptability are to be considered as two poles of a continuum where the target text can be positioned based on its linguistic features examined vis-à-vis the source text and comparable original texts produced in the target language (Toury 1995/2012:79–85). Equivalence is not conceived as an a priori notion that is based on an absolute criterion of adherence to the source text. This means that in a descriptive study the researcher will always assume that equivalence exists. What s/he will unveil is the actual way in which it is realized in terms of the balance between invariance and transformation. This type of equivalence in turn constitutes a steppingstone for discovering the concept of translation that informs the target texts examined, this being defined in terms of the acceptability-adequacy continuum.

At each stage of this process of gradual discovery of facts about the nature of translation and translating, hypotheses are formulated on the basis of empirical descriptions and then verified through further procedures that are applied to an expanding corpus to achieve higher and higher levels of generalization. The procedures elaborated by Toury are largely compatible with the Data-Driven Learning approach developed by Johns (1991a, b) in foreign language education.

This method, as we discussed earlier, is also employed in corpus-based translator training, particularly in the teaching of Languages for Special Purpose (LSPs). The notion of ‘discovery’ plays an important role in Toury’s and Johns’ methodologies. They both require that students and researchers alike progress from empirical data to generalization. The basic corpus-based procedure adopted by Johns is to “Identify—Classify—Generalise” (Johns 1991a: 4) the lexico-grammatical features associated with words that are particularly problematic for advanced learners. The main tool for carrying out this analysis consists of KWIC concordance lines and the procedural steps are those proposed by Sinclair (2003: xvi–xvii).

Conclusion

The agenda we have set out for corpus studies of translation now and in the foreseeable future is to stimulate a continuous process of mutual exchange between theory, research, and practice to reengage the pure and applied branches of the discipline. This entails undertaking research that is firmly grounded in theory and is driven by concerns arising in practice. It also means promoting practice that is underpinned by the insights given by research and elaborating theory that is substantiated and refined by empirical findings. This mutual exchange will promote the interaction between academics and professionals (Pym 2001) and go a long way to resolving the traditional dichotomy between ‘the academic and the vocational in translator education’ (Kearns 2008).