Corpus-based translation studies 15 years on: Theory, findings, applications

Sara Laviosa

Università degli Studi di Bari “Aldo Moro” and Università degli Studi di Roma “Tor Vergata”

Summary

In the field of scholarship known as Translation Studies, computerized corpora have come of age in descriptive as well as applied research. The aim of this paper is to show two main features that characterize corpus-based translation studies at the turn of the century: empiricism and interdisciplinarity. To this end, I draw on Andrew Chesterman’s (2004a, 2007) proposed framework for the similarity analysis of a translation profile. Firstly, I appraise the divergent similarity among descriptive studies of translation universals. Secondly, I introduce a study of Anglicisms whose aim is threefold: to unveil the translation-specific lexical primings of English loan words in the Italian language of business, finance and economics vis-à-vis donor and receptor languages; and to infer the norms that govern the translation of Anglicisms vis-à-vis original text production in a specific domain and genre.

Divergent similarity in the quest for translation universals

Drawing on Chesterman (2004a, 2007), similarity is intended as the sum of relevant samenesses and differences. Divergent similarity results from the process of moving from one entity to more entities. On the other hand, convergent similarity results from perceiving a relation of similarity between two or more distinct entities. The quest for universals in descriptive studies of translation can be examined in terms of divergent similarity. One research endeavour has in fact given rise to a variety of investigations that are similar as regards: rationale; object of study; research model and methodology; accomplishments; posited causes of universals; and the significance of these studies for the scholarly field as a whole.

The rationale can be inferred from two complementary remarks made by Toury and Chesterman. One is a positive evaluation of the notion of translation universal, which is regarded as “one of the most powerful tools we have had so far for going beyond the individual and the norm-governed” (Toury 2004: 29). The other states that the quest for universal features of translations is one way in which descriptive scholars propose and look for generalizations about translation (Chesterman 2000).

The object of study of this research endeavour is variably defined. According to Baker, universals are linguistic features which typically occur in translated rather than original texts and are thought to be independent of the influence of the specific language pairs involved in the process of translation (Baker 1993: 243). Chesterman defines universals as general regularities or laws about translation which can have explanatory force as regards the occurrence of a given feature in a particular translation if they are supported by extensive empirical evidence (Chesterman 2000: 26, 2004b: 39). In a similar vein, Halverson claims that universals are second-level (or internal) generalizations made over numerous empirical studies, and as such they are “explanatory with respect to individual studies of particular linguistic realizations and/or language pairs” (Halverson 2003: 232). For Toury they are
better regarded as probabilistic laws that explain the relations existing among different cognitive, cross-linguistic and socio-cultural variables that influence a particular translational behaviour or its avoidance (Toury 2004: 15). Moreover, the value of such laws lies on their “explanatory power” rather than their “existence” (Toury 2004: 29). Such a claim downplays, in my view, the ontological status of universals and implies that the notion of universal can be effectively used as a heuristic tool for gaining an insight into the nature of translation as a “translating act” and a “translating event” (Toury 1995, 1998).

As regards the research model adopted, universals are explored by putting forward, operationalizing and testing, through a comparative model of translation, general descriptive hypotheses about the existence of similarities between different types of translation, without disregarding either the differences between them or the uniqueness of each particular case (Chesterman 2004b).

The methodology consists in designing unidirectional and bidirectional parallel corpora to investigate S-universals, which refer to “universal differences between translations and their source texts”, and monolingual comparable corpora to study T-universals, which refer to “universal differences between translations and comparable non-translated texts” (Chesterman 2004b: 39).

Corpus studies of universals have refined previous descriptive research into simplification, explicitation and normalization. They have also provided empirical evidence for other posited universals, such as levelling out, the law of interference and the unique items hypothesis (UIH) (see Laviosa 2009 for an overview).

The causes of universals are to be found in human cognition, the nature of translation as a communicative act and the translator’s awareness of his/her socio-cultural role (Chesterman 2000, 2004b). More specifically, lexical simplification, normalization, levelling out and UIH could be accounted for by the asymmetrical cognitive organization of semantic information (Halverson 2003: 218-219, 2007: 118). The importance of human cognition in accounting for translation universals is underlined by Malmkjær, when she observes that “if the concept of universal is to retain any theoretical bite in our discipline, we would do well to reserve it for use in connection with phenomena such as this [UIH], for which it makes sense to produce a cognitively based explanation” (Malmkjær 2008: 57).

The significance of the quest for universals is highlighted by Chesterman, when he affirms that “[c]orpus-based research into translation universals has been one of the most important methodological advances in Translation Studies during the past decade or so, in that it has encouraged researchers to adopt standard scientific methods of hypothesis generation and testing” (Chesterman 2004b: 46). Moreover, studies of universals have formulated and tested explicit interpretive and descriptive hypotheses derived from general theoretical claims, broadened our knowledge of translation, and suggested explanatory hypotheses that can be properly tested by putting forward explicit predictive hypotheses in future studies. In the light of this, research into universals offers an excellent example of the reciprocal relationship between translation theory and description as well as the unidirectional relationship between the ‘Pure’ branch of Translation Studies and its ‘Applied extensions’, as envisioned by Toury (1995). The insights provided by descriptive research have in fact inspired applied scholars such as Scarpa, Stewart and Kujamaki who have investigated universals within a pedagogic perspective (see Laviosa 2009 for an overview).
Summing up, the study of universals is showing, through empirical and interdisciplinary research, that “translations are texts of a particular, specific kind, which reflect the complex cognitive processes and the particular social contexts from which they arise”. (Mauranen 2008: 45). They sometimes show lexical and structural trace of language contact, while at other times they may under-represent features unique to the TL, over-represent elements that are less common in the TL or display a tendency towards conservative or conventional target language use (ibid.).

Looking to the future, the search for translation universals will benefit from:

- reflecting on and refining the methodology;
- widening the range of languages investigated;
- seeing the full potential inherent in the specificity of translation as a bilingual processing situation and an important form of interlingual and intercultural communication that can contribute significantly to our understanding of other forms of language contact, such as the borrowing of lexical or structural patterns across languages (Bernardini and Zanettin 2004; Mauranen 2008).

Convergent similarity in the study of Anglicisms

Linguistic borrowing from English, a phenomenon linked to the status of the English language as an international lingua franca, has attracted much scholarly attention in recent years, particularly in Europe, where the harmonization of a national and transnational identity is intimately related to issues concerning the promotion of multilingualism and the need for mutual intelligibility. It can be safely affirmed that the general assumption underpinning recent research into Anglicisms is the belief that linguistic borrowing, evident in a wide variety of domain-specific discourses, contributes to the creation of a jargon that is inclusive to the members of a particular professional group and exclusive to the outsiders. This results in an enhanced group identity and, for better or for worse, in a shift in linguistic conventions. It follows that linguistic borrowing fulfils an identity-marking function that is as important as the communicative one. It is further believed that, in the wake of globalization, translation is a mediator of language change induced by English source texts, as a result of the operation of the law of interference, whereby “in translation, phenomena pertaining to the make up of the source text tend to be transferred to the target text” (Toury 1995: 275), particularly when translating from a highly prestigious source language to a less prestigious target language (Toury 1995: 278). Yet, the empirical evidence is far from consistent, since there is considerable variation across target languages, domain-specific discourses, text types, and types of Anglicisms. It is therefore still debatable whether translation plays a significant role in the process of Anglicization of the European languages. Hence, translation, as Mauranen (2005: 45) affirms, cannot be regarded as “the enemy within isolated, self-contained languages”.

Against this background, the aim of my investigation is twofold: a) to become clearer about the extent to which translation protects linguistic specificity and cultural difference by resisting the influence of English in business communication; b) to enable translator trainers to formulate and apply “bridging rules” (Toury 1995: 18) firmly grounded in theory and empirical evidence.

The theoretical framework of this study draws on the convergent similarity among: Michael Hoey’s (2005) theory of lexical priming; studies of Anglicisms (e.g. Görlach 2001, 2003 Anderman and Rogers 2005); Gideon Toury’s (1995, 2004) laws of translational behavior;

Since lexical priming provides a substantial input to the present investigation, it is useful to outline its main features. Lexical priming draws on cognitive psychology and builds upon the work of John Sinclair. It is a lexically-driven theory of language that “would seem to offer a dynamic mechanism for change worthy at least of exploration” (Hoey 2005: 188). In cognitive psychology priming is a technique that allows one to study the manner by which the interconnections in the human word-store known as the mental lexicon are constructed (Rumelhart and Norman 1985: 29; Aitchison 1987: 24). A basic priming study consists in presenting subjects with two strings of letters asking them to decide as rapidly as possible whether each is a word or non-word. Typical pairs of items are: nurse plame; bread butter; bread nurse. If the two words are related, the recognition time is significantly shorter than if they are unrelated. The interpretation of this result is that the reading of the first word preactivates the reader’s attention, facilitating the processing of the related word. This preactivation is known as ‘priming’.

Lexical priming is the process whereby “[a]s a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context” (Hoey 2005: 8). While in cognitive psychology the focus is on the relationship between the priming and the primed (or target) word, in lexical priming the focus is shifted towards what is primed, so that “what is primed to occur [in the presence of a priming word] is seen as shedding light upon the priming item” (Hoey 2005: 8).

Lexical priming offers an explanation for a considerable number of linguistic features through the following claims:

- every word is primed to occur with particular other words; these are its collocates;
- every word is primed to occur with particular semantic sets; these are its semantic associations;
- every word is primed to occur in association with particular pragmatic functions;
- every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations;
- co-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations;
- when a word is polysemous, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses;
- every word is primed for use in one or more grammatical roles; these are its grammatical categories;
- every word is primed to participate in, or avoid, particular types of cohesive relation in a discourse; these are its textual collocations;
- every word is primed to occur in particular semantic relations in the discourse; these are its textual semantic associations;
- every word is primed to occur in, or avoid, certain positions within discourse; these are its textual colligations.

Primings are not fixed, they can change through drifting. A drift in the priming of a word, hence a shift in meaning and/or function, occurs when a word is introduced in an unfamiliar
context or co-text or if the language user chooses to override its current primings in his/her own use. Therefore, a drift may occur intralinguistically, i.e. when a word migrates from one domain to another in the same language, and through language contact, i.e. when the receptor language borrows a word from the donor language or when a word in the source language is translated with an exact equivalent in the target language. Drifts in the priming of a word, occurring for a number of members of a particular community at the same time, provide a mechanism for temporary or permanent language change.

A new corpus-based study of Anglicisms

The present study of lexical borrowing aims to make a contribution to our understanding of the domain-specific and text-type-specific primings of lexical Anglicisms in translational and non-translational business Italian, as well as in English and Italian as donor and receptor languages. It also aims to become clearer about the nature of translation as a posited potential source of drifts. In the following sections I will: a) define the object of study; b) put forward three descriptive hypotheses; c) outline the research model and methodology; and d) report on some preliminary findings.

The object of study consists of:
- the lexical primings of Anglicisms in a composite English-Italian comparable and parallel corpus of articles from The Economist and Economy;
- the norms that govern the production of non-translational texts in the receptor language;
- the norms that govern translators’ choices.

Generally speaking, a norm is “[a] regularity in behaviour, together with the common knowledge about and the mutual expectations concerning the way in which members of a group or community ought to behave in certain types of situation” (Hermans 1999: 163). This central notion has given rise a series of formulations of translational norms: initial norm, preliminary norms, operational norms (Toury 1995: 65-66); expectancy (or product) norms, professional (or process or production) norms (Chesterman 1993: 9, 1997: 64-70); constitutive conventions, regulative conventions (Nord 1991: 100-106).

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). Preliminary norms regard translation policy (that refers to the factors determining the choice of text-types or individual texts to be translated) and those related to the directness of translation (tolerance for translating from mediating languages). Operational norms govern the decisions made during the act of translation itself, i.e. matricial norms and textual-linguistic norms. (Toury 1995: 65-66). Expectancy (or product) norms “are established by receivers’ expectations of what a translation (of a given type) should be like, and what a native text (of a given type) in the target language should be like” (Chesterman 1993: 9). Professional (or process or production) norms regulate the translation process itself (i.e. accountability norm, the communication norm, the relation norm) (Chesterman 1997: 64-70). Constitutive conventions “determine what a particular community accepts as a translation” (Nord 1991: 100), while regulative conventions refer to “the generally accepted forms of handling certain translation problems below the text rank” (Nord 1991: 100).
Three research hypotheses have been put forward:

1. English loan words in Italian are differently primed from the original English words, hence they shift in meaning across donor and receptor languages;
2. the translator has the choice of either preserving the primings of the target language or importing the primings of the source language or a mixture of the two;
3. the norms governing translational texts differ from those concerning non-translational texts of the same type;

The study adopts a comparative research model and involves the design and examination of an English-Italian comparable corpus, an English-Italian bilingual parallel corpus, and a monolingual Italian comparable corpus. Moreover, the study envisages to complement corpus data with information about the extratextual factors of the communicative situation (or translating event), i.e. the sender, intention, audience, medium, place, time, motive, communicative function of the target text (Nord 2005: 42, 158) as well as metatexts, i.e. statements and comments made by translators, editors, publishers and other persons involved in the production of the target text (Toury 1995: 65); prescriptive statements contained in translation manuals and guides to business writing.

The corpus analysis involves identifying first of all the Anglicisms contained in the translational Italian subcorpus, using frequency lists retrieved with the WordList facility provided by Wordsmith Tools. The primings of each Anglicism are then examined through the analysis of KWIC concordance lines produced by the Concord facility provided by Wordsmith Tools. The last stage consists in assessing the divergent similarity between: a) these primings and those contained in the non-translational Italian subcorpus; b) the primings in the translational Italian subcorpus and those in the original English subcorpus so as to identify the translators’ choices (using ParaConc); c) the primings in the non-translational Italian subcorpus and those in the original English subcorpus.

Before I move on to report on the initial findings, I will give some background information about the pedagogic context in which this study was carried out. The corpus-based analysis was performed in collaboration with the students attending a 60 credit postgraduate course in specialized translation from English to Italian, which was taught at the University of Bari “Aldo Moro” during the 2008-2009 academic year. As part of the module devoted to the language of business and economics I designed a teaching unit on the translation of polysemic lexical Anglicisms. The teaching objectives were: a) to become familiar with corpora as one of the computer-aided translation tools and resources available to the professional translator; b) to discover the initial and textual-linguistic norms underlying the translation of polysemic lexical Anglicisms; c) to discover Italian native equivalents of well-established English loan words.

The sources of data consist of: a corpus of 71 translated and non-translated comparable articles taken from the Italian weekly magazine Economy and a corpus of 71 English articles from The Economist and their translations in Economy. In the first phase, the lexical Anglicisms contained in the translational subcorpus were identified. By Anglicism it is intended “a word or idiom that is recognizably English in its form (spelling, pronunciation, morphology or at least one the three), but is accepted as an item of the vocabulary of the receptor language” (Görlach 2003: 1). The most frequent Anglicism was business, a well-established English loan word, having been introduced in the Italian lexicon in 1895.
Out of nearly 60,000 running words, 37 occurrences of *business* were retrieved. The analysis of the KWIC concordance lines revealed five discrete meanings of *business*:

I. the work of producing or buying and selling goods or services for money.
II. a high profile area of business where more than one company operates.
III. a) a highly profitable business activity undertaken by a company;  
   b) investment, deal or transaction made by a company.
IV. a large organization that provides services, or that makes or sells goods.
V. volume of business.

For each of the above meanings the collocational, colligational and semantic primings were identified as follows.

I. *Business* occurs with words that refer to other human activities, (*turismo e business*), the geographical place where business is carried out, and the people of different nationalities that are in business; it forms multi-word units (*business hub, area business, segmento business, aree consumer e business*).

II. *Business* occurs with nouns that identify a particular business sector and the position gained in the market, nouns referring to the major players that operate in or impact on it, adjectives describing its qualities, such as diversity, profitability or importance.

III. *Business* occurs with words that refer to the company undertaking a particular business activity and to the type of activity undertaken, verbs referring to the changes undergone by a business, adjectives and nouns describing its main features, such as novelty, solidity or volatility; it forms one compound (*core business*).

IV. *Business* occurs with words referring to the people owning or running a company or to the way in which a company organizes its activities; it forms multi-word-units (*business model, modello di business, business manager, business partner*).

V. *Business* occurs with words that refer to the monetary value (or turnover) of a company or business sector:
   *Al 73enne Ecclestone è rimasto il 25% del gruppo che gestisce il business da 800 milioni di dollari.*  
   [73 year old Ecclestone still owns 25% of the group that runs the 800 million dollar business].

   *Per contrastare il cambiamento del clima anche gli ambientalisti riscoprono il nucleare. Un business da 125 miliardi di dollari.*  
   [To counteract climate change, even environmentalists are rediscovering nuclear power. A 125 billion dollar business’].

Next, business was examined in the comparable corpus of non-translated articles. The number of occurrences was nearly double, 74 against 34. The meanings identified were the same, except for meanings II and V, which also referred to illegal business activities. The analysis of the KWIC concordance lines revealed samenesses and differences (the latter are highlighted in bold):
I. *Business* occurs with words that refer to other spheres of human activity (*business & società, sport e business, musica & business, business & genetica*), the geographical place where business is carried out, and the people of different nationalities that are in business; it forms multi-word-units (*business information, aree di business, clienti business, utenti business*). *Business is used in creative collocations*: Il business resta in porto ‘business stays in the port’. Titanic del business ‘the Titanic of business’. *Il business non è l’unico quadrante su cui far girare le lancette della vita* ‘Business isn’t the only dial on which the hands of life turn’.

II. *Business* occurs with nouns that identify a particular sector and the position gained in the market, nouns referring to the major players that operate in or impact on it, adjectives describing its qualities, such as diversity, profitability or importance. *Business forms compounds* (*business travel, social business, business online*). Business refers to illegal sectors (*il business dei falsi ‘the business of counterfeits’; i business si chiamano droga, prostituzione, racket ‘drugs, prostitution and racketes are known as businesses’*). *Business is used in creative collocations* (*un business che si chiama sconto ‘a business that is called discount’; un business duro come il teak ‘a business as hard as teak’*).

III. *Business* occurs with words that refer to the company undertaking a particular business activity and the type of activity undertaken, verbs referring to the changes undergone by a business activity, adjectives and nouns describing its main features, such as novelty, as well as importance, competitiveness, credibility or profitability; *it forms various compounds* (*core business, business continuity, business case, business plan*), *it strongly collocates with*: possibilità, opportunità, occasioni, fare ‘do’; *it is used in creative collocations and puns* (*il business in una cannuccia ‘business in a straw’; ora faccio business col cuore ‘now I’ll do business with my heart’; un Tornado di business ‘a business tornado’; ho più di un business per capello ‘I’ve got more than one business in my hair’*).

IV. *Business* occurs with words referring to the people owning or running a company or to the way in which a company organizes its activities; it forms multi-word-units (*business unit, unità di business, business development manager*). *Business is used in creative collocations and puns* (*un business fatto di nuvole ‘a business made of clouds’; l’Enav e quel business che è caduto dal cielo ‘Enav and the business that fell from the sky’; il business lievita alla luce del sole ‘business rises in the light of the sun’*).

V. *Business* occurs with words that refer to the monetary value (or turnover) of a company or business sector; *it is used in creative collocations*: *Quel business da 2.5 milioni di sacchi di caffè ‘That 2.5 million sacks of coffee business’; it refers to illegal activities*: *Il business [dei falsi] vale almeno 7 miliardi di euro all’anno ‘the business [of counterfeits] is worth at least 7 billion euros a year’; Cibo Nostro. La Mafia nell’alimentare. Quasi 20 miliardi di incassi per la criminalità organizzata: tanto vale oggi il business mafioso nell’agroalimentare, nelle sue varie declinazioni ‘Our Food. The Mafia in the food sector. Almost 20 billion euros worth of takings for organised crime: that’s how much the Mafia business is worth in the agriculture and food sectors in its various ramifications’*. 

SYNAPS 24(2010)
The results obtained so far show that the collocations, semantic associations and colligations of one sense of the polysemous word *business* differ from those of its other senses. Moreover, the collocations, semantic associations and colligations of *business* in translational and non-translational Italian appear to be divergently similar. Finally, the translations seem to have resisted the influence of English to a degree by limiting the use of *business*.

The next phase involved mapping the Italian target texts onto the English source texts and revealed the following native Italian equivalents of *business* for each of its five meanings:

I) *il mondo degli affari, gli affari, affari, l’attività, attività commerciali*
II) *il settore, l’industria, le industrie*
III) *un’attività commerciale, l’attività, le attività delle aziende*
IV) *un’azienda.*
V) *generating the business → cedendo i prestiti ‘relinquishing loans’.*

The initial norm that can be inferred from these findings is a tendency towards acceptability in the target language, as shown by the absence of *business* in creative collocations or puns in translational Italian. The textual-linguistic norm appears to reveal a preference for native Italian equivalents. There is also one example (see meaning V above) where the original expression containing the word *business* is translated with an equivalent expression which explicates the original sense.

The final phase of the analysis, which was carried out with the subcorpus of original English articles vis-à-vis the subcorpus of comparable non-translated Italian articles, involved the identification of the drifts in the primings of *business* across the donor and the receptor language. Unlike in English, *business* was found to have pejorative overtones in Italian when it conveyed meanings II and V. While in English *business* was found to refer to a small or medium enterprise, in Italian it was used to refer to a large company (meaning IV). In Italian, *business* was sometimes used in creative expressions which usually appeared in the article title or subtitle, thus fulfilling the pragmatic function of attention-getting devices.

**Conclusions**

At the end of this brief journey into the state of the art of corpus-based translation studies I hope I have been able to show that descriptive corpus studies of translation recognize that “[d]escription is not enough. It has to serve a purpose, such as explanation.” (Hermans 1999: 102). Conversely, applied corpus studies of translation recognize that “[t]hey are not intended to account either for possibilities and likelihoods or for facts of actual behaviour, but rather to set norms in a more or less conscious way.” (Toury 1995: 19).

**Acknowledgements**

I wish to acknowledge the collaboration of the students attending the *Corso di Perfezionamento in Traduzione* at the University of Bari ”Aldo Moro” during the 2008-2009 academic year. I am also grateful to Richard Braithwaite for carrying out the translations of the Italian citations retrieved from the translational and non-translation subcorpora.
References


Sara Laviosa