Computerized Corpora and the Future of Translation Studies

Maria Tymoczko

L’approche basée sur le corpus
The Corpus-based Approach
Volume 43, numéro 4, décembre 1998

Résumé de l'article
Cet article traite de la "centralité" des études basées sur le corpus par rapport au domaine entier de la traductologie. L’auteur met le lecteur en garde contre la tentation de faire de la rigueur scientifique une fin en soi par des études quantitatives vides et inutiles.

Citer cet article
COMPUTERIZED CORPORA AND THE FUTURE OF TRANSLATION STUDIES

MARIA TYMOCZKO
University of Massachusetts, Amherst, USA

Résumé
Cet article traite de la «centralité» des études basées sur le corpus par rapport au domaine entier de la traductologie. L’auteur met le lecteur en garde contre la tentation de faire de la rigueur scientifique une fin en soi par des études quantitatives vides et inutiles.

Abstract
This article provides a discussion of the centrality of corpus-based studies within the entire discipline of translation studies. The author warns against the possible danger of pursuing scientific rigour as an end in itself through empty and unnecessary quantitative investigations.

The information age has brought an explosion in the quantity and quality of information we are expected to master. This, along with the development of electronic modes for storing, retrieving, and manipulating that information, means that any discipline wishing to sustain itself in the twenty-first century must adapt its content and methods. Corpus translation studies is central to the way that Translation Studies as a discipline will remain vital and move forward. The essays collected in this volume show that corpus translation studies has the characteristics typical of contemporary emerging modes of knowing and investigating. Corpus translation studies enable us, for example, to encode in compact and efficient forms, to access and interrogate vast quantities of data — more data than any single human being could ever manage to gather or examine in a productive lifetime without electronic assistance. Moreover, the approach allows for and promotes the construction of information fields that suit a new international, multicultural intellectualism, providing for the inclusion of data from small and large populations, from minority as well as majority languages and cultures. Translation corpora make it possible for decentralized, multilocal investigations to proceed thanks to virtually instantaneous access to shared primary materials. Corpora in translation studies lend themselves to joint intellectual endeavors unimpeded by time or space, facilitated by intercommunication across the globe. They permit the reversibility of perspective, the decentering of power. And like large databases in the sciences, corpora will become a legacy of the present to the future, enabling future research to build upon that of the present. Thus, corpus translation studies change in a qualitative as well as a quantitative way both the content and the methods of the discipline of Translation Studies, in a way that fits with the modes of the information age.

Corpus translation studies (CTS) has emerged at a critical time in the discipline of Translation Studies. Growing out of corpus linguistics and thus inherently having an allegiance to linguistic approaches to translation, CTS at the same time marks a turn away from prescriptive approaches to translation toward descriptive approaches, approaches developed by scholars over the last thirty years, notably by polysystems the-
orists such as Itamar Even-Zohar, Gideon Toury, and André Lefevere. CTS focuses on both the process of translation and the product of translation (cf. Holmes 1988: 67; Bassnett 1991: 7), and takes into account the smallest details of the text chosen by the individual translator, as well as the largest cultural patterns both internal and external to the text. It builds upon the studies of scholars working in the descriptive approach to Translation Studies, and those of scholars who have worked with corpora themselves, albeit corpora that have been manually assembled, examined, and analyzed. The essays collected here illustrate that, although the materials of corpora are based upon the language medium of translations, interrogation of corpora can nonetheless serve to address not simply questions of language or linguistics, but also questions of culture, ideology, and literary criticism. Modes of interrogation — as well as care in the encoding of metatextual information about translations and texts — allow researchers to move from text-based questions to context-based questions. The flexibility and adaptability of corpora, as well as the openness of the construction of corpora underlie the strength of the approach.

Inevitably, as is to be expected, the debates of Translation Studies as a whole are mirrored in CTS. For some CTS writers, the notion of "a tertium comparationis" is central, for others the concept has implicitly been superseded (cf. Snell-Hornby 1990). Some writers see norms as one of the most pressing factors to be investigated, while for others the question of norms plays little or no role. Equivalence — a concept that to Jacobson (1959: 233) was "the cardinal problem of language and the pivotal concern of linguistics", but that more recently has fallen into discredit (Snell-Hornby 1988: 13; Van den Broeck 1978) — is still an important part of their discourse for some CTS scholars. The relationship of interpretation to Translation Studies, both theoretically and pragmatically, is addressed within CTS as it is elsewhere in the discipline. The types of literary questions addressed by the "manipulation" school of Translation Studies (see Hermans 1985; cf. Snell-Hornby 1988: 22) motivate some CTS investigations, while literature as such is excluded from other corpora altogether and hence from research done with those corpora. For some CTS investigations, literary texts are seen as optimal, offering the most comprehensive information, while for others literary texts are seen as unrepresentative of natural language. The tendency toward segmentation of pedagogic, pragmatic, and theoretical studies is found in CTS, as it is in Translation Studies and intellectual endeavors altogether.

One of the most important debates in intellectual domains has to do with the search for laws, both practical and theoretical. This debate is also conducted in Translation Studies and mirrored in CTS. The interest in laws — promoted in Translation Studies by Even-Zohar and Toury, among others — follows the tradition of empirical research, behind which lies the presupposition of Western rationalism that science should be in the business of discovering natural laws and that "scientific" results have more value than others. In part a legacy of positivism, these predispositions are intimately connected with a tendency to polarize objectivity vs. subjectivity and privilege the former. A number of CTS scholars promote and justify corpus-based approaches on the grounds that such studies will uncover and establish universal laws of translation.

By contrast, throughout the twentieth century the foundations of such objectivist research were repeatedly called into question. From the work of Einstein and Gödel and Heisenberg in physics and mathematics during the earliest decades of the twentieth century, to mid-century reconsiderations of objectivist premises about literature and history, to still later explorations of subjectivity in the social sciences, modern thought has increasingly come to understand the way that the perspective of the observer or the researcher is encoded in every investigation and impinges upon the object and the
results of study. Such challenges to empiricist claims about objectivity and such assertions of subjectivity — and the large intellectual trajectory behind them — are echoed in some of the approaches represented in this collection and are germane to the development of CTS.

In my view, the value of corpora in translation and of a CTS approach to translation theory and practice does not rest on the claim to "objectivity" and the somewhat worn philosophical tradition claims of this type presuppose. Indeed, as a number of the authors here remind us, behind the establishment of corpora, as behind the design of any experiment or research program or survey, lie intuition and human judgement. Corpora in translation studies are products of human minds, of actual human beings, and, thus, inevitably reflect the views, presuppositions, and limitations of those human beings. Moreover, the scholars designing studies utilizing corpora are people operating in a particular time and place, working within a specific ideological and intellectual context. Thus, as with any scientific or humanistic area of research, the questions asked in CTS will inevitably determine the results obtained and the structure of the databases will determine what conclusions can be drawn. In that sense then, corpora are again to be seen as products of human sensibility, connected with human interests and self-interests.

All the more reason, therefore, to consider the objects of study, the data gathered in the databases, and the parameters defining the corpora themselves very carefully. Central to this concern is the definition of translation, as Sandra Halverson argues. The nature and definition of the category "translation" have been notoriously contentious within the discourses of translation theory and practice. One man's translation is another man's adaptation. The favorite translations of one era are repudiated altogether by another. Like the category game discussed by Wittgenstein in his Philosophical Investigations, there is no simple core identity for the cluster of objects identified as translations by the many societies of human time and space. Rather, translations like games form a category linked by many partial and overlapping "family resemblances", as Wittgenstein called such commonalities. Ultimately the inability to define translation within bounded and finite characteristics and still include in the definition all the objects that human societies have identified as translations led Toury to define translation in a pragmatic way as "any target language text which is presented or regarded as such within the target system itself, on whatever grounds" (Toury 1982: 27; cf. Toury 1980: 14, 37, 43-45). The family resemblances linking translations, like those linking games, have depended on the manifold factors of human culture, including, for example, characteristics of the culture's language, the conditions of "text" production, the role (if any) of literacy, the material culture, economics, social customs, social hierarchies, values, and so forth. In the case of translation, moreover, the processes and products of translation will be correlated not simply with the conditions of one culture, but with those of at least two cultures in interface. It is the very variety of human cultures that leads to the variety of translations and games, and it is that variety that prevents us both pragmatically and theoretically from drawing a neat line, as in classical set theory, around the category translation.4

Although the theory of prototypes may ultimately offer some help in understanding how human beings learn and recognize and remember categories, including cluster concepts such as those of game and translation linked by loose family resemblances, prototypes do not necessarily offer a solution to the theoretical problem of constructing corpora that can be interrogated to reveal translation laws, should a scholar undertake such an exercise. Contrary to Halverson, I take the view that prototypes — including the proposed prototype of the professional translator5 — are themselves culturally
defined and culture bound, rather than universal. As Eleanor Rosch, the pioneer in prototype research, specifically stipulates, the theory of prototypes is intended to address issues in categorization that "have to do with explaining the categories found in a culture and coded by the language of that culture at a particular point in time" (1978: 28); moreover, she acknowledges that "many categories may be culturally relative" (1975: 193). Thus, a prototype is hardly the point of departure for a search for general or universal laws of translation, for one would then immediately have to ask compromising questions about the results. For example, what language, what place, and what time were privileged in the selection of the prototype? Our own? If so, we consign our theoretical work to hopeless ethnocentrism and so doom it from the outset.

To take as an analogue the well-discussed category chair, although we might discover on the basis of research done with contemporary American subjects that the prototype of chair is a wooden, four-legged desk chair, the modern prototype is not necessarily a guide to the prototypical chair of other times and other cultures. In the case of chair, for example, there is good reason to believe that in the Renaissance and earlier, the prototypical chair would have been very different from the configuration of the modern prototype. In view of material evidence from the period, the prototype would perhaps have been a three-legged apurtenance, with a triangular seat and a relatively small back, chairs such as are common from the Renaissance period and earlier. Four-legged chairs are only practical if floors are level, whereas three-legged chairs, like three-legged stools, are more useful, more common, and, thus, presumably, more prototypical in cultures where floors are made of mud or uneven paving stones, as floors were in most houses in pre-modern conditions. Indeed such three-legged stools and chairs — if chairs and stools are used at all — continue to be the norm (and, hence, probably the prototype) in most peasant cultures today.

The problems of beginning with a prototype in a search for general laws governing a cluster concept like translation are even more complex than those posed by implicit ethnocentrism. In the case of the would-be prototype of the professional translator suggested by Halverson, for example, we might go on to ask how one defines or picks out a professional translator now, leaving aside the question of the past or other cultures. Is a professional translator a translator trained in a translation school or program? (Even if that person never works as a translator or never is paid for translation?) At what point does the non-professional student translator make the transition to professional? (If a student translator publishes a translation, is that piece by a professional?) Or is a professional translator a translator whose work has been published or otherwise remunerated — hence an a posteriori determination which will include in many cases amateur translators who have "made good" or "become successful"? If one could resolve questions such as these to the satisfaction of all, the question would still remain as to when the prototypical translator is translating prototypically, i.e. under what conditions professional translation is deemed to exist or to have taken place. Thus, are all translations of a professional translator prototypical, or only those produced in some stipulated professional context? Moreover, are all types of texts translated by a professional translator prototypical, or only a limited subset? And so forth. The recursive nature of the stipulation required for defining a prototype in the case of the concept like translation illustrates the difficulty that follows from using prototypes as a foundation for the search for general laws of translation.

Such issues can also be illustrated by questions related to Wittgenstein's category of game. What is the prototype of game in the modern period? The prototype of an abstract cluster concept like game is much harder to establish than that of a material category like chair. There may in fact not be any example of game that functions as a
prototype for an entire culture as complex and differentiated as contemporary Western cultures are. It may be that there are several competing prototypes of game, each particular to a specific subculture, for example, football, tennis, cards, video, and so forth. And again, as with categories like chair, a conceptual category like game or translation also changes through time. One might surmise, for example, on the basis of textual, artistic, and archaeological evidence, that the prototypical game for many people in medieval France was dice gaming — a pastime that is currently unlikely to be prototypical of game for very many people, if for anyone, in Europe or America. And once more the problem of establishing a prototype becomes recursive, for even if one could make allowances in research for multiple as well as ethnically decentered prototypes — say dice gaming as the medieval European prototype of game — what are the prototypes that are themselves in turn presupposed by the typical instance of game at any one cultural locus? In the case of the example under discussion, one should then ask, for instance, what is the prototype of dice. Almost invariably for a modern informant the term will suggest a pair of cubical objects (although other types of dice exist, as any player of Dungeons and Dragons can testify), while in medieval Europe oblong dice, made of bone and used in threes, were much more common and, thus, presumably more prototypical.

These are some of the difficulties with basing the construction of translation corpora on a theory of prototypes. As Halverson proposes, translation corpora based on modern Western "prototypes" might be developed and interrogated so as to reveal general characteristics of contemporary translations done in circumstances and cultural settings congruent with the examples underlying the corpora. Such corpora will be enormously useful and valuable, but they will not yield general laws of translation applying to all times and all places and all languages. To discover general laws of translation, if indeed such laws exist, a question that Wittgenstein's arguments about categories should lead us to consider very seriously, at the very least what will be needed are corpora representing as many types of translations as are known from the whole of human history. It may be, however, that such a quest is a positivist chimera, the commonalities so restricted for a category like translation that the effort is unlikely to provide the field of Translation Studies with much information of lasting value or transferable worth and, therefore, would not be worth the effort.

If the primary purpose of CTS is neither to be objective nor to uncover universal laws, what is the specific strength of descriptive studies of translation facilitated by computerized corpora that can be electronically searched and manipulated? Over and over again the essays collected in this volume speak to this question, offering many divergent and exciting new directions for Translation Studies. The multiple possibilities that can be foreseen even at this embryonic stage of CTS are welcome, suggesting that corpus translation studies will be an open rather than a convergent approach to the theory and practice of translation. One can project the construction of many different corpora for specialized, multifarious purposes, making room for the interests, inquiries, and perspectives of a diverse world.

Comparison is always implicit or explicit in inquiries about translation, and there is often a tendency to focus on likeness rather than difference and to rest content with perceptions of similarity. Whether focused on linguistic or literary or cultural matters, comparative disciplines — among which Translation Studies takes its place (cf. Bassnett 1993) — all have a predilection for focusing on similarity or sameness, as Dwight Bolinger, writing about linguistic studies, observes (1977: 5):

Always one's first impulse, on encountering two highly similar things, is to ignore their differences in order to get them into a system of relationships where they can be stored,
retrieved, and otherwise made manageable. The sin consists in stopping there. And also in creating an apparatus that depends on the signs of absolute equality and absolute inequality, and uses the latter only when the unlikeness that it represents is so gross that it bowls you over.

This impulse, which partly drives the interest in laws of translation, could be a projection of the development of a narrow form of CTS, particularly at the beginning of the exploration of this approach to corpora of translations. Bolinger correctly sounds a cautionary note: like Translation Studies as a whole, CTS must find ways to move beyond systems of relationship that focus on likeness and to avoid being locked into a binary apparatus that acknowledges only perceptions of absolute equality and inequality.

One reason that CTS is likely to avoid the pitfall of fixation on similarity and remain open to difference, differentiation, and particularity, is the sheer variety in natural languages themselves and the multiplicity of theoretical and practical consequences resulting from the manifold language pairings possible in translation. This “infinite variety”, which CTS is more able to apprehend and include in its purview than traditional methods of Translation Studies, militates against universalist programs of research and universalist conclusions. From a positivist point of view, such variety and divergence might be seen in a negative light, but from most other perspectives, the potential of CTS 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 in translation constitutes the chief appeal of this new approach to Translation Studies. Particularly at this juncture of history, we need to know the specifics of different languages in translation, the individual particularities of specific pairings of languages in translation exchanges, and the characteristics of translation as cultural interface at different times and places and under different cultural conditions. Such differences teach us as much as or more than commonalities of translation, and they contribute as much or more not only to our theoretical investigations of translation phenomena but to the practical concerns of situating translation in a decentered, multicultural world.

For a long time in the history of translation theory and practice, difference was perceived in a negative way, as a departure from fidelity, a sign of the loss inherent in the translation process. In very different ways, both Eugene Nida’s school of “dynamic-equivalence” translation and more recent approaches to translation — from those of Philip Lewis (1985) to the feminist translation theorists (Bassnett 1992, 1993: ch. 7) to the Brazilian school promoting “cannibalism” (Vieira 1994) — have valorized difference in translation. It is clear that CTS has the potential to be one means of challenging hegemonic, culture-bound views of texts, translations, and cultural transfer. It is a powerful tool for perceiving difference and for valorizing difference as well.

What are the next steps in the development of corpus translation studies? As we have seen, there are many leading edges in this field, just as there are in the field of Translation Studies as a whole. Nonetheless, looking at the most exciting new work in Translation Studies, one notes certain common themes and common commitments. Thus, for example, there is an growing commitment to integrate linguistic approaches and cultural-studies approaches to translation, to explore their reciprocal relationship, thus turning away from the invidious competition and isolation that plagued these two branches of the discipline for some time (cf. Baker 1996). Second, the leading work in Translation Studies shows an ever more sophisticated awareness of the role of ideology as it affects text, context, and translation, and as it affects the theory, practice, and pedagogy of translation as well. And, finally, as in other disciplines, the pioneering work in
Translation Studies involves adapting modern technologies to the discipline's needs and purposes. The essays at hand illustrate that corpus translation studies is dead center on all of these developments, once again suggesting the leading role it will play in the discipline of Translation Studies in the coming decades.

In building for the future, CTS must take care not to diminish itself, falling into the fetishistic search for quantification that plagues many "scientific studies" and makes them ridiculous, empty exercises. Researchers using CTS tools and methods must avoid the temptation to remain safe, exploiting corpora and powerful electronic capabilities merely to prove the obvious or give confirming quantification where none is really needed, in short, to engage in the type of exercise that after much expense of time and money ascertains what common sense knew anyway. As the dour Vermonter might put it, "Do bears sleep in the woods?" Researchers must take care to ask "the right questions": to pose questions and construct research programs that have as their goal substantive investigations that are worthy of the powerful means deployed by CTS.

In building for the future, CTS must also recognize the imperatives and honour the claims of historicism. Within Translation Studies there have now been many case studies related to historical poetics, for example, amassing considerable data on the wide variation of the role of translation in culture and the range of the norms of translation practice. The gains of such work must be incorporated into the design and construction of corpora. Researchers must of course avoid the obvious trap already discussed of being locked into the translation norms of the present, and of presupposing such norms in the construction of corpora. But beyond that obvious weakness, ontological and epistemological commitments in the design of corpora must include dedication to the past and to other cultures as well.

The development of corpora and CTS methods represents a long-term investment for the field of Translation Studies. As we now set the foundation of a legacy that will come to fruition in the future, it is important to begin to envision the widest possible range of corpora that can be established, the uses to which corpora can be put, of questions that can be asked solely of corpora, and of methods that can be utilized with corpora for both theoretical and practical results. From envisioning these things we must proceed to encourage their development and realization.

Finally, CTS is once again 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. One of the most encouraging aspects of the pioneering studies of CTS is the way that seemingly technical and theoretical interrogations come to have practical potential and immediate applicability, not only for the teaching of translation but for the work of the practising translator as well.

Notes
1. See T. Tymoczko (1979) for an example of the way that electronic capabilities change disciplines in a qualitative manner. The results of corpus translation studies may at times generate a scepticism similar to that which computer proofs initially raised among mathematicians.
2. See for example the case studies in Even-Zohar (1990); Lefevere and Jackson (1982); and Hermans (1985). See also Brisset (1996).
3. That is, to establish criteria that will individually be necessary and collectively be sufficient to pick out all translations and only translations.
4. For a general discussion of the philosophical issues related to categories see Wittgenstein (1953); Rosch (1978); Gleitman, Armstrong, and Gleitman (1983); and Lakoff (1987).
5. As a word of caution, one should note that a prototype is not defined by researchers a priori, but is derived from actual empirical research with actual subjects speaking a specific language. It is quite possible that research would show that a professional translator is not prototypical of the concept translator to most speakers of English.
6. It should be noted that Rosch’s conclusions about prototypes are based primarily on research having to do with basic biological objects and man-made objects that can be easily visualized and unambiguously represented with line drawings (Rosch 1975, 1978), though she has also done work with categories such as colors and forms (Rosch 1973), categories that can likewise be represented visually. Rosch repeatedly offers disclaimers about extending the results to abstract categories and other complex cases such as translation, for which little research having to do with prototypes currently exists.

7. Note the additional problem that we know of types of translation that we have no data for and, hence, cannot include in corpora—for example, oral translation in preliterate, prehistorical conditions.

REFERENCES


SNELL-HORNBY, Mary (1990): "Linguistic Transcoding or Cultural Transfer? A Critique of Translation Theory in Germany", Translation, History and Culture, Susan Bassnett and André Lefèvre (Eds), London, Pinter, pp. 79-86.

TOURY, Gideon (1980): In Search of a Theory of Translation, Tel Aviv, Porter Institute for Poetics and Semiotics.


