Document generated on 04/24/2024 12:08 a.m.

Meta
Journal des traducteurs
Translators' Journal

META

Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical/Descriptive Categories and a Conception of the Object of Study

Sandra Halverson

Volume 43, Number 4, décembre 1998

L'approche basée sur le corpus
The Corpus-based Approach

URI: https://id.erudit.org/iderudit/003000ar
DOI: https://doi.org/10.7202/003000ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Halverson, S. (1998). Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical/Descriptive Categories and a Conception of the Object of Study. *Meta*, *43*(4), 494–514. https://doi.org/10.7202/003000ar

Article abstract

This paper discusses the issue of representativeness in the creation of general translation corpora. In the course of a in-depth examination of the stages involved in the selection of texts which adequately represent the target population, it demonstrates that prototype categories are better suited than the all-or-none classical ones to reconcile existing theoretical statements about what constitutes legitimate data in translation studies with the new methodology developed by the corpus-based approach.

Tous droits réservés © Les Presses de l'Université de Montréal, 1998

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

TRANSLATION STUDIES AND REPRESENTATIVE CORPORA: ESTABLISHING LINKS BETWEEN TRANSLATION CORPORA, THEORETICAL/DESCRIPTIVE CATEGORIES AND A CONCEPTION OF THE OBJECT OF STUDY

SANDRA HALVERSON
University of Bergen, Bergen, Norway

Résumé

Cet article traite de la représentativité dans la création de corpus de traduction générale. En examinant en profondeur les étapes de sélection des textes qui décrivent le mieux la population cible, l'auteur démontre que les catégories prototypiques sont mieux adaptées que les catégories classiques plus rigides pour la réconciliation des assertions théoriques existantes sur ce qui constitue des données valables en traductologie avec la nouvelle méthodologie créée par l'approche basée sur le corpus.

Abstract

This paper discusses the issue of representativeness in the creation of general translation corpora. In the course of a in-depth examination of the stages involved in the selection of texts which adequately represent the target population, it demonstrates that prototype categories are better suited than the all-or-none classical ones to reconcile existing theoretical statements about what constitutes legitimate data in translation studies with the new methodology developed by the corpus-based approach.

1. GENERAL REMARKS

All academic disciplines strive for some degree of consensus about the types of theory they develop, and the data and methods used to test them. Indeed, such a consensus is often taken as the defining criterion of a "mature science" (Kuhn 1970). While total or ultimate agreement is neither possible nor, indeed, desirable, much of what has been termed "normal science" arguably proceeds under precisely such conditions (Kuhn 1970). Concerns of this nature have recently been expressed in several publications in the field of translation studies, as many translation scholars address the need to consider questions of methodology and data in the field (see Gile 1991; Toury 1985, 1995; van Doorslaer 1995; Fraser 1996). Indeed, the collection of papers in this volume may constitute a major step towards establishing the position of translation-relevant corpora[1] as data, and methods adapted from corpus linguistics as appropriate means of studying translation. Achieving such an end would imply the enrichment of the discipline as a whole in the direction of new means of testing existing theories, and new types of theoretical questions.

However, none of these gains are absolute; neither are they divorceable from more fundamental philosophical questions. At the present stage of development, there seems to be little discussion of the epistemological implications of adopting one methodology or set of data over, or in addition to, another. And methods, data and theories are, as we know from current philosophies of science, inextricably linked. Consequently, a science which addresses questions of methodology, data, and theory without a basis in the metatheoretical question of how it conceives of its object is not sufficiently grounded. This paper is an attempt to alleviate the problem, if ever so slightly, by discussing the relationships between these various levels of scientific enquiry in the study of translation. Thus, the focus in what follows is on the relationships between various conceptions of the object of enquiry, the theories derived to explain that object, and the data and methods used to test and refine those theories. The immediate concern is to consider the more fundamental relationships implied and required by the use of corpora in the investigation of translation.

As it incorporates current work with relevant corpora, i.e. learner corpora, comparable corpora, etc. (see Baker 1993; Granger 1994), translation studies is already being provided with a number of new theoretical questions, most notably the set of "universal features of translation" put forward by Baker (1993; see also Toury 1995 on "universal laws of translation"). Furthermore, several studies have already used corpus-based approaches to address various aspects of that particular theoretical issue (see Laviosa-Braithwaite 1996, 1997; Laviosa 1998; Øverås 1995). As work on implementing the various techniques of corpus analysis proceeds using variously designed corpora, it becomes imperative that we establish the means to compare and integrate this work. Part of the groundwork for ensuring comparability of results lies in careful and rigorous selection of the criteria used for the construction of the corpora and the means used in the analysis. This, in turn, entails a clear articulation of the basis on which construction criteria are selected, which must be the conception of the object of study. In other words, all discussion of corpus text selection and classification, the types of analysis adopted, and the significance of the findings must be grounded in an explicit description of what the enquiry takes to be its object.

The articulation of what a field of enquiry takes to be its object must be provided by theories within that field. In what follows, I shall argue that current theories of translation do not provide the kind of articulation that is needed to accommodate the demands of corpus techniques and the statistical analyses they foster.[2] Furthermore, I contend that this problem may be resolved by viewing "translation" as a prototype category. I will present the arguments underlying this proposition, starting with a description of the underlying epistemological assumptions on which corpus linguistics and descriptive translation studies are based. The next step will be to consider those assumptions as they relate to relevant theories, to which empirical studies must turn for hypotheses, descriptive categories, etc. The vital link between these two areas of concern and the question of the object of study is provided by the concept of representativeness. After discussing the question of representativeness with reference to translation theory and data, I shall argue that representative corpora may be created for the study of translation on the basis of an understanding of the object category as a prototype one.

2. REPRESENTATIVENESS: THE LINK BETWEEN OBJECT, THEORY AND DATA

For the purpose of this discussion, it is important to keep in mind that a corpus-based methodology is one which incorporates emphasis on empirical investigation and

quantitative analysis (however, see note 2). As a consequence, the adoption of such a methodology implies conformity to standards that are widely understood to constitute a scientific method (see Leech 1992). In brief, if corpus-based methods are to be used to study translation, then the corpora must be constructed in accordance with the standards and measures that apply to such an empirical tool, and the analyses carried out using these tools must do the same. Perhaps the most fundamental issue in the use of any collection of data is representativeness, or, in technical terms, the extent to which the data may be said to include "the full range of variability in a population" (Biber 1993: 243). In what follows, the question of representativeness will be outlined within the empirical, statistical paradigm implied by corpus linguistics.

2.1. General Remarks on Representativeness

In a highly informative article on the issue of representativeness, Biber (1993) discusses common concerns in ensuring that a collection of data (e.g. a corpus) represents its object of enquiry. He states (1993: 243, my emphasis):

> Typically researchers focus on sample size as the most important consideration in achieving representativeness: how many texts must be included in the corpus, and how many words per text sample. Books on sampling theory, however, emphasize that sample size is not the most important consideration in selecting a representative sample; rather, **a thorough definition of the target population** and decisions concerning the method of sampling are prior considerations.

Thus, a prerequisite to the construction of a corpus is a careful description of what that corpus is intended to represent. Within the statistical paradigm in which the article is written, Biber continues to outline exactly what a definition of a target population entails (1993: 243):

> Definition of the target population has at least two aspects: 1) the boundaries of the population — what texts are included and excluded from the population; 2) hierarchical organization within the population — what text categories are included in the population, and what are their definitions.

The significance of defining the target population is paramount. As Biber emphasizes, without a clear definition it is impossible to assess whether the corpus is sufficiently related to the object of enquiry or ultimately adequate for the purposes for which it is intended. He warns (1993: 243):

> In designing text corpora, these concerns [those outlined above, SH] are often not given sufficient attention, and samples are collected without a prior definition of the target population. As a result, there is no possible way to evaluate the adequacy or representativeness of such a corpus (because there is no well-defined conception of what the sample is intended to represent).

Thus, in the interests of making corpora both representative and adequate for a research purpose, one of the first steps (within the methodological paradigm in which we are operating) is to clearly define the target population. In the following section, we shall see how this has been done in the process of constructing existing monolingual corpora. Vital to our understanding of this process is a clear picture of the role theory plays in that definition.

2.2. Definition of a Target Population

The process of defining a target population consists of specifying the boundaries and internal structure of the category (Biber 1993). In specifying the boundaries, the

objective is to devise a means of determining which instances are to be included in the category, and which are not. Outlining the internal structure of the category implies specifying a means of further classifying the category at large. The only means we have of establishing the parameters needed to define a target population is a theory of our object of enquiry. As Biber states, "theoretical research should be prior to corpus design" (1993: 243). But prior even to the theoretical search for a motivated delimitation of the object is consideration of the purpose for which the corpus is to be used. Sinclair distinguishes between specific and general purposes, the first of which he defines as "those where the end result is the only important issue" (1992: 379-380). An example of such a specific purpose might be the creation of a spell checker. General purposes, Sinclair suggests, include much more complex tasks, such as studying conditions for machine translation or human-computer communication. Also among the general purposes is the study of "language [...] for its own sake," as well as "the application of pre-existing descriptions to corpora, to test and improve them" (1992: 380). It is this most general level that is of concern here.

Given that the purpose of constructing a corpus is to study a language and to test given descriptions (or theories), the first step is to designate the boundaries of what may be included as examples of that language. This is most often done for monolingual corpora with reference to the concepts of "naturally-occurring" language, and "native speakers." Quirk (1992: 459) even restricts this latter concept to "mature native speakers" in his description of the principles applied in the construction of the Survey of English Usage in the early 1960s. Thus, there seem to be two criteria applied in determining the boundaries of the object category: a situational one (naturally-occurring) and a demographic one (native-speakers). Furthermore, there is apparently a considerable consensus on the theoretical justification of such a boundary.

As for the second part of the target population definition, the specification of internal structure or hierarchies, Biber builds on previous theoretical research on register analysis (see Halliday *et al.* 1964). His focus is on the situational criteria designated by the theory (primary channel, format, setting, etc.), though he also allows for the possibility of a set of demographic criteria (lines of occupation, sex, age, etc.) as well. Moreover, Biber (1988) is able to argue the validity of his situational parameters on the basis of prior statistical analyses. For monolingual corpora in "western societies" (1993: 244), his proposal is built on theoretical categories which have already been exposed to considerable empirical/statistical analysis.

Another approach to structuring the target population is the categorization used in the design of the British National Corpus.[3] The classification system for written texts makes use of three parameters: domain, medium and time. The latter two are relatively straightforward and specify the type of publishing medium, e.g. periodical, book, and the publishing date, respectively. The domain parameter, on the other hand, is less unambiguously assigned. Texts are classified first according to whether they are informative or imaginative, and the former are then subclassified in terms of subject area, or domain, in technical terms. Examples of such domains include arts, belief and thought, commerce, leisure, etc. The basis for this classification system is similar to Biber's, i.e. prior consensus on the theoretical and/or empirical validity of the categories. For the purpose at hand, however, the key idea is that for a general purpose monolingual corpus both the boundaries of the target population (the object category) and the internal structure have been provided by theoretical consensus, and subsequently, empirical testing.[4]

2.3. Specifying a Sampling Frame

In a discussion of the representativeness of a corpus, it is necessary to distinguish between the target population, as outlined in the preceding sections, and a sampling frame, which is "an operational definition of the population, an itemized listing of population members from which a representative sample can be chosen" (Biber 1993: 244). For monolingual written corpora, there are several pre-existing bibliographical sources which may serve as sampling frames. In the construction of the British National Corpus, for example, texts were selected from a variety of information sources, including catalogs of books published per annum, lists of books in print, bestseller lists, library lending statistics, lists of current magazines and periodicals, periodical circulation figures (BNC Users' Manual: electronic version). The decisions made about the use of these various data sources must, once again, be made on the basis of theoretical notions pertaining to the relevance of various types of text, and the relative significance of the different types. It must also be mentioned that practical concerns such as time and money quite often factor into the decision as well. Key to our discussion, however, is the distinction between a target population, which is a theoretically derived construct, and the sampling frame, which is an operationalization of that construct.

The relationship between these various concepts may be represented schematically, as in Figure 1. An academic discipline takes a portion of the world as its object of enquiry (and indeed justifies its existence on the basis of arguments aimed at proving the existence of that object as an independent phenomenon). For the purpose of corpus construction, the conception of the object which a discipline more or less agrees on provides the motivation for defining a target population. The specification of a target population then provides the basis for selecting a sampling frame. The sample taken from that frame constitutes the corpus, which is thus a representation of a larger set of phenomena.

Figure 1 also outlines the mutual relationships between decisions made at various levels. A decision regarding the type of object to be focussed on is determined by the type of ontological and epistemological commitments that prevail. In other words, it depends on what kind of things one thinks can exist, and how one believes it is possible to learn about them. The decisions made in defining a target population occur at the theoretical level; that is, they are based on, and/or determined by, theoretical commitments derived from the ontological and epistemological commitments which are prior to them. In other words, to define a target population, one needs a theory. And in order to create a theory, one needs an idea of what can be a theory, and what kind of things one may construct theories about. Furthermore, the decision to define a target population in the first place is a methodological one, which is related to theoretical assumptions at the same level. If, according to one's theory, one's ideas about theories in general and the world in general, a reasonable thing to do is to construct a corpus, then one must define a target population in a specific way. Finally, the specification of a sampling frame and selection procedures are purely methodological decisions which derive solely from the methodology selected. In brief, if one is to define a target population on the basis of all of the prior decisions outlined here, then there are certain procedures which must be followed.

It must be emphasized that the process is not strictly linear, though it may seem so in the explication given here. Though the various decisions are chronologically and logically linear if viewed from the perspective of one isolated corpus constructing task, it is also true that insight gained at one level of enquiry will have an effect on the other levels as well (as indicated by the two-way and circular arrows in Figure 1). Thus linearity must be seen as the result of the current point of view, and as such represents merely a frozen picture of a dynamic process or system.

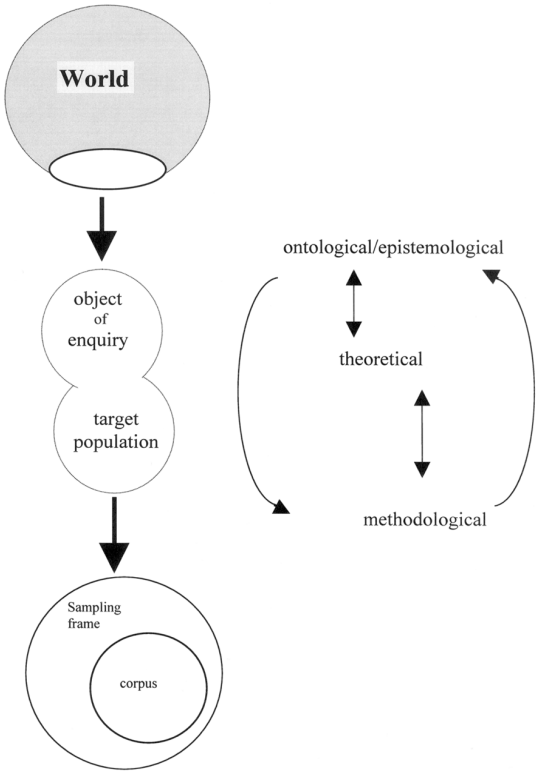

Figure 1
Relationships

3. REPRESENTATIVENESS IN TRANSLATIONAL CORPORA: DEFINING THE TARGET POPULATION

The basic premise for the following argument is that we are considering a general purpose translation corpus, one at the same level of generality as the most general monolingual corpora referred to above. In other words, for the sake of this argument, it is to be assumed that the purpose for constructing a translational corpus is to study translation in its own right, and to test extant descriptions and theoretical constructs against a collection of relevant data. However, the question of the type of analysis comes more immediately to the fore in the case of translational corpora than in the case of monolingual corpora, as there seem to be two alternative, or perhaps two complementary, forms of translation corpus being constructed today. On the one hand, there is the type that includes translated texts and their sources, either from one language into another or both ways. Such a structure is used in the "core corpus" of the English/Norwegian translation corpus described in Hofland & Johansson (1994), and Johansson *et al.* (1996). On the other hand, there are corpora of translated texts and "comparable" texts, for the most part texts created in situations/contexts which are similar to the translations' contexts (Baker 1993, 1995; Aijmer *et al.* 1996; Laviosa-Braithwaite 1996, 1997; Laviosa 1998). Some corpora combine the two types. In the

Norwegian corpus mentioned above, for example, the "core corpus" is supplemented by comparable texts in both languages. Thus, for translational corpora, a decision regarding the value of various types of comparisons is often built into the design itself. The basis for making that decision is, of course, theoretical preference.

The issue of representativeness under discussion here, however, is not concerned with the type of comparison that might be carried out. As a result, the question of comparable versus parallel corpora of various types will not be addressed (see Granger 1994 and Baker 1995 for discussion of varying corpus types and differences in terminology). The question of representativeness is concerned with the translations themselves, and is irrespective of the type of comparison to be carried out (if any). What is at stake here is what the **translations** in a corpus are taken to represent.

3.1. Drawing the Boundaries

As mentioned in section 2.2, defining the target population consists of at least two tasks: specifying the boundaries and describing the internal structure of the category (the population). Furthermore, as was also pointed out earlier, the means of arriving at both category boundaries and internal structure are provided by current theories. Accordingly, the aim of this section is to investigate what translation theory has to offer in this respect.

In monolingual corpora, the question of what constitutes a text in a language has been resolved with reference to demographic and situational criteria, as mentioned in section 2.2. Unfortunately, translation theory does not provide us with an analogous set of criteria. In what follows, we shall first consider what options translation studies does provide for a definition of the object of enquiry. Secondly, we shall consider why adopting criteria analogous to those used in linguistics at large is not feasible.

Over the past twenty-odd years, there have been two main ways of answering the question of what the concept of translation may be said to include. These two approaches differ significantly in their underlying philosophical assumptions, and consequently in their approach to the object of enquiry (see Halverson 1996b, 1997 for a more detailed discussion of the philosophical differences between the two). In brief, and indeed rather superficially, it may be said that the two represent an *a priori* and an *a posteriori* approach (for detailed discussion see Toury 1980, 1995; Delabastita 1991). In the former, "translation" is defined, usually in terms of a range of equivalence types, prior to empirical study. Accordingly, texts included for study are those which are considered to fulfill the requirements specified in advance (see e.g. Wilss 1982; Koller 1989).[5] The latter approach takes a broader view, and considers any translation that is "presented or regarded as such within the target culture, on whatever grounds" (Toury 1985: 20) to be legitimate data. It is the latter view which seems to be most generally adopted in the field today, in spite of its rather problematic implications (see Pym 1995; Halverson 1997). Given such a definition, it follows that any translation which meets the necessary and sufficient condition stipulated by Toury is equally valuable in theoretical terms (see Chesterman 1993). Consequently, any translation which is "presented or regarded as such" would be included in the target population and would thus be a viable candidate for ultimate selection in a general purpose corpus.

Unfortunately, such a conclusion is at odds with another theoretical assumption in the field, namely the privileged status of professional translations. The primacy of professional translations has been established within the framework of a norm-based theory of translation (see Toury 1980, 1995). A logical consequence of such a theory is that norms will be most clearly and frequently manifested in the behaviour of a specific group of individuals. As outlined by Chesterman, this group must consist of "compe-

tent professional translators" (1993: 7). At the same time, there seems to be a rather resilient minority who believe that other types of translations are also theoretically interesting, i.e. those referred to as "natural translations" (see Harris 1977a, 1977b, 1980, 1988, 1992; Toury 1995). Such translations are produced under rather different circumstances, "everyday" ones, and also by a different set of individuals, people with "no special training" in translation (Harris 1973). Given Toury's definition above, however, it would be hard to argue against including such translations as equal members of our category of translations. As long as such texts are identified as translations in the target culture, they are full-fledged members, and legitimate data.

One might also address the issue of non-professional, e.g. student, translations in the object category. This issue is highly relevant, as ample empirical evidence suggests that the processes involved in professional as opposed to non-professional translation are quite different (Jääskeläinen & Tirkkonen-Condit 1991; Jääskeläinen 1989, 1990, 1993; Tirkkonen-Condit 1988, 1990, 1992; Séguinot 1991; Krings 1987). More methodological aspects of this question have also been addressed, here with an aim to limit the use of student translations as data in translation studies (Gile 1991). However, given that such texts are definitely identified as translations in a target culture, it is logically difficult to exclude them from a definition of a target population.

This description of the *status quo* in translation theory presents us with a seemingly irreparable situation. On the one hand, there seems to be at least some agreement on an *a posteriori* definition of the object. Such an approach entails that all translations which are recognized as such are equally interesting for the formation of a general theory of translation. On the other hand, there seems to be a consensus that professional translations are more theoretically significant than non-professional ones, and more central than translation carried out in situations characterized by the term "natural translation." Thus translation theory seems to be at odds with itself.[6]

Let us then turn to the criteria which have seemingly served monolingual corpus designers so well. The first criterion applied in the construction of general purpose monolingual corpora was the demographic one which limited the object of study to language production by native speakers. The application of this criterion results in a uniform set of texts in the sense that any text (spoken or written) is equally valuable as an example of the language in question (at least with reference to the text production parameter). In other words, there is no reason to assume that one person's, or group of people's, language is more theoretically interesting than an other's.

In translation studies, this demographic equality is not evident. There are no "native translators." Indeed, it has been argued that the concept of a "native translator" may be posited (Toury 1984, 1995). Arguments for the existence of such individuals are developmental: it has been suggested that translators evolve through a series of stages from a "predisposition" to translate (which is considered "coextensive with bilingualism") to the emergence of the skill through processes of socialization (Toury 1995: 241-258). According to Toury, the emergence of the skill is "coextensive with the ability to establish similarities and differences *across* languages" (1995: 248). Thus, such a process of development would involve activating a second kind of ability (which Toury terms "interlingualism"), that is ontologically distinct from the first. In other words, the development of a translator is considered to involve two different types of ability: linguistic aptitude in two languages, and the comparative ability described above. It is at this critical point that the analogy with the native speaker breaks down. The development of bilingual individuals into what most people would refer to as translators requires a certain development of the mediation faculty, and this evolution is not at all natural or in any way comparable for all bilingual persons (see Beardsmore 1982).

Native speakers, on the other hand, all develop from a predisposition for language[7] to master the norms of their native language. This is, of course, a circular argument, as any given language is defined by what native speakers produce. However, it is precisely the inability of translation studies to agree that translation is what all translators do (or that what all translators do qualifies as "translation") that has brought us to the present impasse in defining our object of study.

It should also be pointed out that even if one was to assume that translators are socialized in analogous ways to native speakers of a given language, the actual conditions of that socialization would indicate that such a comparison might warrant caution. Within the confines of a behavioural approach to translation, one might point out that factors influencing the acquisition of translating behaviour include, first of all, the potential for self-selection: people most often choose to become translators or to take foreign language courses, and their motivations for doing so may be multiple. Secondly, the socialization of would-be translators often involves a different kind of setting (an institutional one, for the most part) and a different set of rewards and sanctions than those involved in learning a first or second language (for example, the possibility of a job or the threat of unemployment).

In sum, the concept of "native translator," while fruitful in the attempt to outline processes of development, is not, for our purposes, sufficiently analogous to the concept of native speaker. Having said that, it must also be emphasized that the translation activity encompassed by developmental studies, as well as studies of so-called "natural translation" (see Harris 1988 for an annotated bibliography) also have a place in the field of translation studies. The argument being presented here is not that such translations should not be included in our concept of the object of study, it is that they should not be included **on the same grounds** or **with the same status** that the language production of native speakers is included in the study of a given language.

A further, and indeed complicating, factor in demographically determining the status of text producers is the question of directionality. There is a general consensus on the primacy of translations done into the translator's first language (Baker 1992, 1995). At the same time, this view has come under attack from various fronts. Malmkjær (1993) argues in favour of translation **out of** the first language in the interest of qualitative differences (measured in levels of equivalence), and Marmaridou (1996) argues the same point on the basis of metaphoric mapping. At the very least this disagreement implies that the direction of the translation vis-à-vis the translator's first language must be specified, and function as a situational parameter (see section 3.2).

The second criterion which applies to the definition of a target population for monolingual corpora (in addition to the native speaker concept) is the situational specification that language must be "naturally-occurring." At first glance, such a criterion seems to be relatively unproblematic for designers of translation corpora. It must be noted, however, that translations do not occur "naturally" in the same way that other types of text do. This may seem rather obvious, but if such a criterion is to be applied, then the notion of "naturally occurring" translations must be articulated by translation theorists. To my knowledge that remains to be done. Does publication ensure that a translation has come into existence under natural circumstances? Does the existence of a commission guarantee naturalness? The latter seems to be one of the most obvious candidates for inclusion in a list of parameters describing naturalness in translation (though it is explicitly linked to the set of translations carried out by professionals), and indeed, such a condition seems implicit in scopos theory (i.e. the aim/purpose (the scopos) of a translation, which determines its final composition, is "negotiated with the client who commissions the action" (Vermeer 1989: 174)). Further work is needed here.

In the preceding paragraphs, an attempt has been made to consider how the boundaries of a target population may be drawn in the interest of constructing a general purpose translation corpus. First, translation theory was investigated to search for relevant guidelines. The unfortunate consequence of that analysis is that translation theory is currently in conflict with itself. Current approaches to defining the object of study do not provide any means of recognizing the primacy of certain types of data. At the same time, theorists like to operate on the assumption that certain types of data are more central than others, and empirical work has shown that, in terms of text producer (demographic) criteria, translations vary significantly. Secondly, in an attempt to create criteria which are analogous to those applied in the construction of general-purpose monolingual corpora, it became clear that the demographic concept of "native speaker" is not directly applicable in translation studies, and that the situational specification of "natural" situations, while promising, remains to be elaborated. At present, translation theory does not provide us with an adequate means of determining where the boundaries of a target population might be drawn.

3.2. Internal Structure

In section 2.2 it was pointed out that the second constituent of the target population definition was a specified means of structuring the category internally. Such structure is provided through the specification of so-called strata, or subcategories. Biber (1993) provides a list of parameters which may be applied in the construction of strata. Furthermore, he distinguishes between parameters which may be used in defining a sampling frame, and parameters which "must be represented in a corpus but [which] do not provide well-defined sampling frames" (1993: 246). Biber's parameters are all situational, and include such factors as channel, format, setting, addressor, addressee, etc. The strata used in the written portion of the BNC were mentioned in section 2.3. These included parameters which were both situational (functional) and content-related.

Thus, it appears that the subcategories currently most used all seem to be derived from various parameters of language function, use or content, or a combination of all three. Once again, it must be emphasized that the means applied to structure a set of texts have been supplied by theoretical consensus, and, in the case of Biber's work, have been tried and tested empirically. In the issue of translation corpora, the question to be asked is twofold, and directly parallel to the question considered in section 3.1: has translation theory provided the means to structure the object category, or does linguistics at large provide us with adequate parameters?

Considerable theoretical work has been focussed on the derivation of translation-relevant text type categories. Driving this work is the recognition that various features of texts cluster together, and that these characteristics have an effect on the type of translating activity carried out. The translational relevance of text type was probably first associated with the work of Katharina Reiss (1969, 1976, 1981, 1989).

Reiss provided a functional typology, which posited a set of three "pure" text types: informative, expressive and operative. These were derived on the basis of a distinction between various "communicative forms" (1981: 124), "communication of content," "communication of artistically organized content," and "communication of content with a persuasive character," respectively. Reiss also pointed out that texts rarely, if ever, exist in their "pure" forms, and that most texts consist of mixes of the above types.

Notably, however, in later work Reiss also included what she referred to as "special cases" (1981: 131), which are translation situations in which "there is a difference between the original text function and the function of the translation" (1981: 131).

Interestingly, later theorists of the functional school (Reiss & Vermeer 1984; Nord 1991) maintain that such functional differences are the norm, rather than the exception. The relevant issue for our purposes, however, is that the existence of such differences, regardless of how frequently they are found, necessitates the derivation of a **translational text typology**: as long as there is a possibility that source and target functions do not coincide, then source text function is not an adequate means of classification. In other words, as the subject of our enquiry is **translated** texts, and not texts in general, the parameters used to define strata must operate at the level of translation, rather than text production in general. In specifying translation functions, for instance, Reiss suggests text "reproduction" for research purposes, or communication of content for "a certain further use" (1981: 131).

From a different perspective, the inadequacy of one non-translational text classification has also been noted by Aijmer *et al.* (1996). In the construction of their Swedish/English comparable corpus, Aijmer *et al.* found that a generic, or "stylistic," stratification system similar to the one used in the Brown Corpus[8] was problematic as "only certain text categories are translated" and because "it is difficult to achieve a good genre match between text samples from the two languages" (1996: 81). The result for this research group is a classification system with "obvious gaps" which result from the realities of translating activity in Sweden. This realization provides support for the theoretical conclusion reached above, and as such, the push for translationally relevant strata comes from both theory and practice. The problem is quite simple. As Reiss puts it, "The criteria are not to be derived from the question 'to what end and for whom has the text been *written?*,' but from the question 'to what end and for whom is the text *translated*'" (1981: 131).

In sum, it seems that in the question of how to structure the translation category internally, translation theory has a bit more to offer. The task of specifying functional parameters is well underway, and the task remaining for a corpus constructor is to determine which of these may be adapted to the specification of a sampling frame, and which are unsuitable for that purpose.

In turning to the question of adapting parameters for corpus stratification from those used by monolingual corpus constructors, it becomes obvious that it is impossible to separate the interests of linguists and translation scholars as clearly here as it was in the consideration of theoretical means of drawing boundaries for the target population. In other words, the parameters proposed by Biber, for example, and those used in the construction of the BNC build on the same types of criteria which underlie the typologies derived for translated texts. All of these classification systems include parameters having to do with text function and the situation in which the text was produced or is intended to be received. Thus, there is no question that functional, or content-related, parameters may be used in classifying translations. They are already being used in this way. At the same time, however, it is clear that the parameters used to stratify monolingual corpora are derived on the basis of a desire to segregate **texts**, not **translations**. Thus, while these parameters are useful, they are not adequate.

To reiterate: stratification of a translation corpus must derive from parameters that are valid for the categorization of translations, not of texts in general. Whether or not such a stratification also makes use of strata found in monolingual corpora is incidental, and is a function of whatever similarity may exist in text-creating situations in general. What exactly such translational parameters might consist of remains to be fully articulated. Translation theorists have already argued quite convincingly that the function of the translation is one such parameter. Furthermore, the possibility of a change in function between the translation and its source, as posited by the scopos/functional-

ist theoreticians (Reiss & Vermeer 1984; Nord 1991), suggests that functional constancy might also be included as a parameter. The theoretical debate on the role of directionality (see section 3.1.1) suggests that this too must serve as a parameter.

In conclusion, even after this brief perusal of translation theory, we might agree that the task of structuring a translational corpus is not insurmountable, and that the theoretical means of doing so are at hand (in the form of functional, situational parameters). This may seem odd, given the difficulties that emerged in our attempt to draw the boundary of the target population. Intuitively, it may seem strange that we are able to devise means of classifying something which we are unable to delimit. This conclusion, however, is not surprising if we consider our attempts at categorization in light of theories which aim to account for categorization itself. In the following section, theories of categorization will be investigated so that we may better understand why our efforts to define a target population have met with so little success. The proposal which will be put forward in the following section is an attempt to resolve the theoretical inconsistencies which prevent us from solving the definition problem.

4. PROTOTYPE VERSUS CLASSICAL CATEGORIES: IMPLICATIONS FOR BOUNDARY DRAWING AND INTERNAL STRUCTURE

Before even discussing a solution, it may be helpful to remember that the main problem at hand is how to ensure that translational corpora are representative of what we want them to represent. At the risk of stating the obvious, let us recall that what we want corpora of any type to represent is the object of study, and representativeness is what allows for generalization from the particular.

In the preceding sections, I have argued that translation theory to date presents conflicting views on what may be included or excluded from the object category, and at the same time seems to suggest that certain types of translation are more theoretically significant than others. I have also argued that two of the criteria most frequently used in the construction of monolingual corpora are not immediately applicable to the task of designing a translational corpus. In discussing the internal structuring of the object category, I suggested that several of the parameters used in constructing monolingual corpora may be applicable for translational corpora, but that these would not be adequate in and of themselves: additional parameters to describe translational features are also needed.

Implicit in the discussion in the first three sections, and indeed explicit in the terminology used, is the idea that an object of study may best be definitionally viewed as a category. This view of a definition implies that the task of defining is, in fact, the same as determining which instances of an object may be included in the concept. It should be obvious that this approach is highly relevant to the task of deciding which types of text to include in a corpus. As suggested at the end of section 3, I believe that considering our attempts to define our object in light of contending views of how we categorize will provided us with the means to solve the definition problem.

4.1. Categorization

The main distinction to be made in the context of this discussion is between so-called "classical" and "non-classical" categories. Understanding this distinction is fundamental to the argument, as the theoretical implications of moving from the one to the other are of considerable magnitude.

In one of the major works on categorization, George Lakoff introduces classical categories as follows (1987: 6):

> From the time of Aristotle to the later work of Wittgenstein, categories were thought [to] be well understood and unproblematic. They were assumed to be abstract containers, with things either inside or outside the category. Things were assumed to be in the same category if and only if they had certain properties in common. And the properties they had in common were taken as defining the category.

Classical, or Aristotelian, categories are thus categories which may be defined in terms of necessary and sufficient conditions. These conditions are considered to be the **essential** properties that all members of the category have in common. As Lakoff states above, these may be used to define the category. In line with this reasoning, concepts (categories) are defined with reference to the necessary and sufficient conditions which constitute a fixed boundary for admission: instances which do not fulfill the conditions are excluded from the category, and instances that qualify become full-fledged members. As a consequence, all members of the concept category are equal. It must also be noted that, in accordance with the classical notion, the essential features of any concept reside in the world as such, and are not humanly determined. In other words, the distinctions between categories that are created by different concepts are **to be discovered**; they are not man-made.

The challenge to classical categories came from two directions, philosophy and psychology. The philosophical arguments were presented in the later work of Wittgenstein (1958), who presented several arguments against classical categorization. For our purposes, three main lines of reasoning are particularly relevant. The first of the three was the argument against the idea that all category members share a set of common features. Wittgenstein's argument, using the example of the game category, was that members of a category may be related through a series of *family resemblances*. In other words, certain members of a category may share some characteristics, while others share different ones. There may be no one characteristic that all of the members have in common. In his most well known example, the game category, Wittgenstein outlines family resemblances as follows (1958: 32):

> Consider for example the proceedings that we call "games." I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? [...] Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball games, much that is common is retained, but much is lost. — Are they all "amusing"? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck: and at the difference between skill in chess and skill in tennis. [...] And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail.

The second line of reasoning presented by Wittgenstein was also illustrated using the game category. Here Wittgenstein showed that the category may be infinitely extended along the dimensions that seem to pertain. His argument was that as long as a new game resembles one or more old games in one or more dimension (entertainment value, skills required, equipment employed, etc.) then it will qualify for inclusion in the category. Thus the fixedness of category boundaries is shown to be illusory.

Wittgenstein's third argument challenged the equality of membership implied by the classical definition of a category. As mentioned above, a category defined by necessary and sufficient conditions will imply equality of membership to all members satisfying the conditions for membership. Using the game category, again, as well as the

"number" category, Wittgenstein argued that there are clear indications that certain members are considered more central, or "better" examples than others. In the number category, for example, it was argued that integers appear to be central, while in the game category, dice may be posited as a less central member.

The challenge from psychology came in the work on categorization done by Eleanor Rosch and her associates (Rosch 1973, 1975a, 1975b, 1977, 1978, 1981). Rosch is credited with developing an actual theory of prototype categories, which has been tested by Rosch herself as well as various others (Ekman *et al.* 1972; Berlin & Kay 1969; Brown 1958, 1965; Berlin 1968; Berlin *et al.* 1974). This research focused on such issues as the structure and acquisition of categories, testing for prototype effects, and searching for sources of prototypes. Much of the early work was done on colour concepts, and these studies indicate that colour concepts are indeed prototypical: there are "better" or "worse" examples of, say, the colour red. Studies also showed that specifying a fixed boundary for the "red" category is impossible. In conclusion, both the philosophical foundations and the empirical "reality" of classical categories have been undermined. It has been shown that for all types of concepts, ranging from the most concrete/physical (the colour red) to the most abstract ("number"), categories defined in terms of necessary and sufficient conditions are inadequate both as regards their empirical "match" with "reality," and their philosophical premises. A theory of prototype categorization offers a much more coherent and promising alternative.

4.2. Prototypical Translation

In an earlier paper (Halverson 1996b), I argued that current approaches to defining "translation" as an object of study make use of a concept defined in the classical way.[9] In that paper, I proposed that adopting a prototype definition of the object concept has enormous potential for resolving a number of current theoretical conflicts and philosophical lacunae. In the following, I intend to pursue that line of reasoning, and show how adopting a prototype approach would provide an operational solution to the definition problem for purposes of constructing a representative corpus.

But first, a brief word on the status of prototype definitions. By suggesting that a prototype definition would constitute an "operational" definition, I seem to be suggesting merely another rather ad hoc attempt to come up with a working definition. As such, my proposal would constitute nothing new. There is, however, a very fundamental difference. In choosing to adopt a prototype category, we do not merely surrender the notion of being able to define our object **now** (which is what most people assume when they adopt a working definition). On the contrary, we surrender the notion that we will **ever** be able to provide an absolute definition. We do not do so without considerable gain, however. In exchange, we are provided with the means to resolve the conflict outlined in section 3, the apparent discrepancy between current attempts to define translation and the desire to privilege certain types of data. In resolving that problem, we are granted the means to define a target population on the basis of a coherent theoretical definition of the object of study.

Classical definitions of "translation" propose categories in which all members are equal. Such a category may be depicted as in Figure 2. In a classical category, the boundaries are fixed and defined by necessary and sufficient conditions. This is shown by the solid line boundary and the even distribution of instances within the category. In a prototypical category, however, the boundary is not fixed, and the members are not equal. Such a category is illustrated in figure 3. Here the dotted line indicates an extendable (and permeable) boundary, and the cluster of instances in the center indicates where there is greatest consensus on the central position of given instances. The

link between the patterns displayed by these two types of categories and the problem of defining "translation" is quite succinctly stated by Harris (1992: 100):

> ... definitions of "translation" are affected by the kind of translation their authors are interested in; and though they must all agree on a few basics — two languages, transposition between them — otherwise the very concept of "translation" would be shattered, they vary in extension and in the relative importance accorded to the component features.

Harris is, in fact, stating what is probably more or less the case with definitions in many fields of academic enquiry. But rather than despair at the lack of definitional coordination, it would be wise to consider a means of positioning the various definitions and their extensions relative to each other.

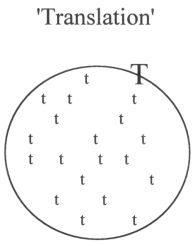

Figure 2
A classical translation category

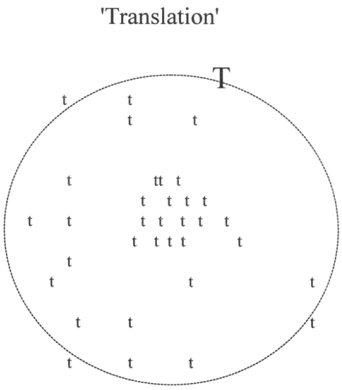

Figure 3
Translation as a prototype category

Before doing so, however, it would be wise to recall why some find definitional disagreement problematic. A lack of coordination can not be maintained in a classical category like that shown in Figure 2. In such a category, there must be agreement on where the boundary should be drawn, as the boundary is taken to correspond unambig-

uously to essential features inherent in the object itself. Definitional disagreement is problematic because someone must ultimately be right.

But what we have seen in sections 2 and 3, and what Harris put so clearly, is that definitions are relative. The subsequent argument is that the logical consequence of definitional relativity is a move to a category such as that in Figure 3, which allows for the positioning of various definitions relative to each other, with a clustering at the center where there is greatest agreement or overlap. The problem of whose definition is right disappears, and the various definitions are all encompassed in an overall concept.

In considering the representations in Figures 2 and 3, it should be clear that using these two types of category will result in very different target populations and very different samples. In a classical category with equal members (translation types), it would be necessary to sample various types of translation equally, which would result in generalizations across potentially very different subcategories.

With a prototypical translation category, one might start by defining the target population as the prototype, which would allow for narrower generalizations, but ones which would be relatable to others in a meaningful way.

4.3. Structuring the Corpus

A sensible first step in the investigation of a prototype category would be to concentrate efforts at the center. In the case of the translation category, this might arguably be constituted by professional translations, at least for the cultures of industrialized, western countries.[10] The parameters used in determining professionalism are by necessity culture-bound, and will thus vary.[11] For a general purpose corpus, the target population could be defined as professional translations into a given language, or from a given language into another given language, depending on whether one wishes to consider questions pertaining to the role of the source language.

A general corpus may also need to cover non-central members of the category. This could be done by constructing ancillary corpora which group other apparent clusters in the overall category. The boundaries of such corpora could be drawn by grouping together texts produced by "natural translators," and/or texts produced by students at translator training institutes, or by students of foreign languages (which would constitute a separate group).

It must be emphasized, however, that the "boundaries" that are being suggested here are to be considered of the permeable, non-definite type suggested by Figure 3. That is, there are sure to be instances which straddle the tenuous border, and ones for which there will be disagreement. Furthermore, the design structure that I have suggested here is substantively different from that of a corpus based on the classical hierarchical classification system, such as that used in most corpora today. The difference is shown in Figures 4 and 5.

Given the broad definition provided by Toury, and the philosophical entailments that follow from it, a corpus would, in order to be representative, have to include equal numbers of all of the types of translation suggested in Figure 4, and would generalize on the basis of overall comparisons between all the texts in the corpus. The stratification would obviously include additional parameters, but the point is to indicate the basis for generalization, and its obvious contradiction of other theoretical views.

Figure 5, on the other hand, presents a better picture of the theoretical status quo in translation studies. At the center are professional translations. At various distances from the center are other clusters, which have been shown to have certain aspects in common, and to be considerably different from each other as well as from the center. A general translation corpus might define its target population as the prototype. One may

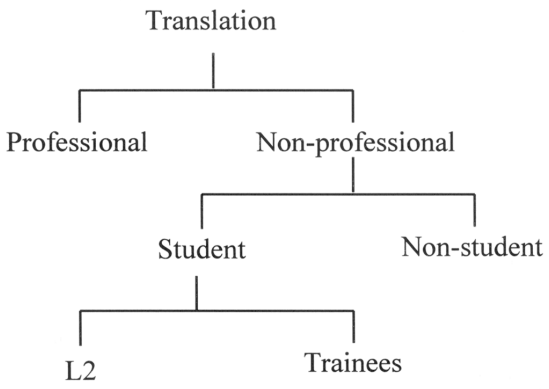

Figure 4
Traditional hierarchical classification (classical category)

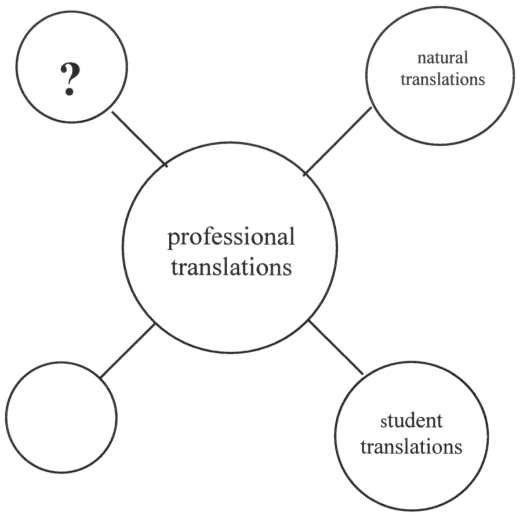

Figure 5
A prototype-based corpus

also choose to define additional populations, which would be kept separate in ancillary corpora. Any generalizations which may result would be on the basis of the prior, and separate, investigations of the center and peripheral "subcategories," and second level comparisons could factor in consideration of relative positioning, i.e. distance from the prototype.

It is also obvious from the above discussion that the kind of analytical apparatus and parameters necessary for the successful adoption of a prototype approach are

nowhere near available. Thus nothing like the kind of stratification proposal, or parameter list, put forward by Biber for monolingual corpora is yet attainable. There is considerable need for both further theoretical work and empirical investigation of the implications of adopting a prototypical approach. This is always the case, however. Preliminary proposals launch tentative investigations, which in turn help to determine which avenues to pursue and which to abandon. As Biber (1993: 256) states, "Theoretical research should always precede the initial corpus design and actual compilation of texts [...] Regardless of the initial design the compilation of a representative corpus should proceed in a cyclical fashion." His cyclical process continues from theoretical research and pilot empirical investigation, though corpus design, compilation of a portion of the corpus, empirical investigation, and then return to corpus design, etc. Thus, the arguments presented in this paper are situated at the initial stages of that cyclical process. However, even as rudimentary a prototype design as that suggested above provides a better means of matching translation theory and translational data than more classical approaches.[12]

5. CONCLUSIONS

The argument presented here is actually quite simple. At its core is the assumption that if we are to pursue a future in translation studies which involves the construction and use of large-scale computerized corpora (and I believe that is one of many very promising routes to take), then the corpora must be constructed in accordance with the rules that apply. The most fundamental rule is that we make sure our collection of text material represents, in a legitimate way, our object of study. Consequently, if current theories of our object of study present us with an object category in which some members seem to be more central, or better examples, than others, then our corpus should reflect that.

It is important to stress the preliminary nature of this realignment in the perception of our object category. Whether or not professional translations are, indeed, as central as theorists seem to agree, or whether other features are more promising in describing the prototype, is an empirical question. The degree to which the prototypes found in different languages or cultures are similar or different must also be studied empirically. What a prototype category structure provides us with is a means of addressing the relativity of definitions, and of studying various kinds of translation within an integrated framework. It also provides us with a means of coping with the quite obvious asymmetries in the category members (the different (types of) translations). In resolving these two major difficulties, we become able to define our target population in a more motivated way, and to suggest means of positioning related data and thus improve the bases for our generalizations.

The most immediate practical consequence of this proposal is that translational corpora would focus on professional translations, and, if desired, include ancillary corpora for the purposes of comparing of periphery to center. Thus the target population definition would be well grounded in theory. Another consequence must be that the parameters used for the purposes of structuring (stratifying) the corpus, would be derived on the basis of translational interests, rather than text production features in general. It has been shown that attempts to stratify translation corpora on the basis of generic or stylistic criteria have not been entirely satisfactory for various reasons. Suggestions as to the types of situational and demographic parameters that might be employed were given in section 3.2.

The main thrust of this paper is, however, to provide some of the grounding necessary for continued work in corpus-based translation studies. The reasoning has been that, in order to move in the suggested direction, the discipline must incorporate consideration

of all levels of enquiry. In closing, then, let us recall the relationships depicted in Figure 1, where the mutual and ever-evolving relationships between the world and scientific attempts at describing it were outlined. In discussing the figure, attention was directed to the mutual influence of ontological, epistemological, theoretical and empirical studies and reflection. It was pointed out that decisions taken at one level build on decisions made on the other levels, and at the same time have effects on or implications for them. The aim of an academic discipline and the individuals working within it is to ensure that their efforts aim at presenting a coherent and unified approach to their project. In short, the objective is to make sure that the data and methods used in empirical research are suitable for testing and constructing theories, which, in turn, are useful in describing and explaining how and why the object of enquiry can be the way it appears to be. And what is taken to be the object of enquiry must be in line with assumptions pertaining to the kinds of things that exist, and the ways we can study and understand them.

The proposal put forward in this paper is one which is motivated both by a consideration of developments at the level of translation theory and at the level of ontology and epistemology. It seems that growing awareness of a new way of looking at the world, and knowing about the world, represents a good match with theoretical attempts to describe one little corner of it like translation. Thus, in arguing for a prototype conception of "translation," I am arguing for a coherent package of philosophical assumptions and theoretical explication, and, in this paper, have suggested what I consider to be the logical implications of that alignment for the selection and organization of relevant data, the final component in the circular process.

Notes

1. For the purposes of this paper, translation(al) corpora are considered to be large-scale computerized corpora, in line with Baker (1995: 226).

2. This is not to imply that corpus analysis is solely quantitative. On the contrary, corpora may also be used for various more or less qualitative descriptions, and indeed the two must be considered complementary. The argument here is that the methodology itself stipulates the use of certain quantitative and statistical approaches in, for example, the construction process (see section 2). In addition, implicit in the argument for the creation of large-scale machine-readable corpora is that they provide a sounder basis for quantitative and statistical analyses which, in turn, allow for more valid generalizations (see Leech 1992; McEnery & Wilson 1996).

3. The British National Corpus is a 100-million-word corpus of contemporary spoken and written English. It is the result of a collaboration between several academic and publishing institutions, and the corpus itself was launched in 1995.

4. For a more detailed discussion of the theoretical foundations of corpus linguistics see Chafe (1992), Fillmore (1992), Leech (1992).

5. In all fairness, it most be noted that Koller has since revised his position. He too, however, is aware of the implications of attempts to define the object of study (see Koller 1995).

6. Chesterman (1993) presents a clear articulation of this problem. He attempts to resolve it within a norm-based approach by arguing that a given set of norms, those demonstrated by professionals, comes to be adopted as a standard for other performers. I have argued elsewhere (Halverson 1996b) that this tack is insufficient, as it fails to account for both the ontological status of so-called "expectancy norms" and the underlying motivation for promoting certain norms over others.

7. It is not the intention to embark on a discussion of what parts of the language faculty, if any, are innate. The argument here is that **if** it is assumed that there is an innate predisposition, then the development of that into a full-fledged skill is different in the case of language and translation.

8. The Brown Corpus (1961) is a corpus of approximately one million words of written American English text.

9. Chesterman presented this argument, with a slightly different focus, in his 1993 paper. He proposed a quite different solution, however, which I argue against in Halverson (1996b and 1997).

10. All prototypes will be culturally determined. The proposal to adopt professional translations is being put forward for a certain group of countries in which the arguments for privileging such data were promoted.

Whatever kind of translation might be central to the "translation" category in other types of cultures remains to be elaborated.

See also Koller (1995) on the development of a concept of translation as a reflection of cultural development.

11. On parameters which may be used in describing professionality and the cultural relativity of their application, see Halverson (1996a).

12. In Halverson (1997), I begin empirical investigations into the relevant parameters of the prototype category. In that paper I test for so-called prototype effects. In another study which is also in progress, I search for evidence of the processes which generate the category structure, in line with the types of categories outlined by Lakoff (1987).

REFERENCES

AIJMER, K., ALTENBERG, B. & M. JOHANSSON (1996): "Text Based Contrastive Studies in English. Presentation of a Project", *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 1994,* Aijmer, Altenberg and Johansson (Eds), Lund, Lund University Press, pp. 73-85.

BAKER, M. (1992): *In Other Words*, London, Routledge.

BAKER, M. (1993): "Corpus Linguistics and Translation Studies: Implications and Applications", *Text and Technology: In Honour of John Sinclair*, Baker, Francis and Tognini-Bonelli (Eds), Amsterdam/ Philadelphia, John Benjamins, pp. 233-250.

BAKER, M. (1995): "Corpora in Translation Studies: An Overview and Suggestions for Future Research", *Target* 7 (2), pp. 223-243.

BEARDSMORE, H. B. (1982): *Bilingualism: Basic Principles*, Philadelphia/Clevedon, Multilingual Matters.

BERLIN, B. (1968): *Tzeltal Numeral Classifiers*, The Hague, Mouton.

BERLIN, B. (1976): The Concept of Rank in Ethnobiological Classification: some Evidence from Aguaruna Folk Botany", *Language, Culture and Cognition: Anthropological Perspectives*, Ronald Casson (Ed.), New York, Macmillan, pp. 92-113.

BERLIN, B. and P. KAY (1969): *Basic Color Terms: Their Universality and Evolution*, Berkeley, University of California Press.

BERLIN, B., BREEDLOVE, D. and P. H. RAVEN (1974): *Principles of Tzeltal Plant Classification*, New York, Academic.

BIBER, D. (1988): *Variation Across Speech and Writing*, Cambridge, Cambridge University Press.

BIBER, D. (1993): "Representativeness in Corpus Design", *Literary and Linguistic Computing*, 8 (4), pp. 243-257.

BROWN, R. (1958): "How Shall a Thing Be Called?", *Psychological Review*, 65, pp. 14-21.

BROWN, R. (1965): *Social Psychology*, New York, Free Press.

CHAFE, W. (1992): "The Importance of Corpus Linguistics to Understanding the Nature of Language", *Directions in Corpus Linguistics,* Svartvik (Ed.), The Hague, Mouton de Gruyter, pp. 79-97.

CHESTERMAN, A. (1993): "From 'is' to 'ought': Laws, Norms and Strategies in Translation Studies", *Target*, 5 (1), pp. 1-20.

DELABASTITA, D. (1991): "A False Opposition in Translation Studies: Theoretical versus/and Historical Approaches", *Target*, 3 (2), pp. 137-152.

DOORSLAER, L. van. (1995): "Quantitative and Qualitative Aspects of Corpus Selection in Translation Studies", *Target*, 7 (2), pp. 245-260.

EKMAN, P., FRIESEN, W. & P. ELLSWORTH (1972): *Emotion in the Human Face*, Elmsford, N. Y., Pergamon Press.

FILLMORE, C. (1992): "'Corpus Linguistics' or 'Computer-aided Armchair Linguistics'", *Directions in Corpus Linguistics,* Svartvik (Ed.), The Hague, Mouton de Gruyter, pp. 35-60.

FRASER, J. (1996): "The Translator Investigated. Learning from Translation Process Analysis", *The Translator*, 2 (1), pp. 65-79.

GILE, D. (1991): "Methodological Aspects of Interpretation (and Translation) Research", *Target*, 3 (2), pp. 153-174.

GRANGER, S. (1994): "From CA to CIA and Back: An Integrated Approach to Computerized Bilingual and Learner Corpora", *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 1994,* Aijmer, Altenberg and Johansson (Eds), Lund, Lund University Press, pp. 37-51.

HALLIDAY, M.A.K., MCINTOSH, A. and P. STREVENS (1964): *The Linguistic Sciences and Language Teaching*, London, Longman.

HALVERSON, S. (1996a): "Data Selection in Descriptive Translation Studies: the Status of Professional Translations", Paper presented at Unity in Diversity? International Translation Studies Conference, Dublin City University, Dublin, 9-11 May 1996.

HALVERSON, S. (1996b): "Conceptual Categories and the Object of Study in Translation Studies: Moving from Classical to Prototype", Paper presented at *Transferre Necess Est,* the Second International Conference on Current Trends in Studies of Translation and Interpreting, 5-7 September 1996, Budapest.

HALVERSON, S. (1997): "The Concept of Equivalence in Translation Studies: Much Ado about Something", *Target*, 9 (2), pp. 207-233.

HALVERSON, S. (work in progress): "Basic-level Semiotic Processes? Searching for Prototype Effects in the 'Translation' Category".

HARRIS, B. (1973): "La traductologie, la traduction naturelle, la traduction automatique et la sémantique", *Cahiers de linguistique*, 10, unpaginated.

HARRIS, B. (1977a): "The Importance of Natural Translation", *Working Papers on Bilingualism*, 12, pp. 96-114.

HARRIS, B. (1977b): "Towards a Science of Translation", *Meta*, 22 (1), pp. 96-114.

HARRIS, B. (1980): "How a Three-year-old Translates", *Patterns of Bilingualism*, Afendras (Ed.), Singapore, RELC, pp. 370-393.

HARRIS, B. (1988): "Natural Translation: Annotated Bibliography", *TRANSST,* 3, pp. 5-7.

HARRIS, B. (1992): "Natural Translation: A Reply to Hans F. Krings", *Target*, 4 (1), pp. 97-103.

HOFLAND, K. and S. JOHANSSON (1994): "Towards an English-Norwegian Parallel Corpus", Fries, Tottie, Schneider (EdS), *Creating and Using English-language Corpora*, Amsterdam/Atlanta, Rodopi, pp. 25-37.

JOHANSSON, S., EBELING, J. and K. HOFLAND (1996): "Coding and Aligning the English-Norwegian Parallel Corpus", *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 1994*, Aijmer, Altenberg and Johansson (Eds), Lund, Lund University Press, pp. 87-112.

JÄÄSKELÄINEN, R. (1989): "Translation Assignment in Professional vs Non-professional Translation: A Think-aloud Protocol Study", Séguinot (Ed.), *The Translation Process,* Toronto, H.G. Publications, School of Translation, York University, pp. 87-98.

JÄÄSKELÄINEN, R. (1990): *Features of Successful Translation Processes: A Think-Aloud Protocol Study*, Lic. Ph. Dissertation, University of Joensuu, Savonlinna School of Translation Studies.

JÄÄSKELÄINEN, R. (1993): "Investigating Translation Strategies", Tirkkonen-Condit and Laffling (Eds), *Recent Trends in Empirical Translation Research (Studies in Languages. University of Joensuu, Faculty of Arts, no. 28)*, Joensuu, pp. 99-120.

JÄÄSKELÄINEN, R. and S. TIRKKONEN-CONDIT (1991): "Automised Processes in Professional vs. Non-professional Translation: A Think-aloud Protocol Study", Tirkkonen-Condit (Ed.), *Empirical Research in Translation and Intercultural Studies*, Selected papers of the TRANSIF seminar, Savonlinna, 1988, Tübingen, Gunter Narr, pp. 89-110.

KOLLER, W. (1989): "Equivalence in Translation Theory", Chesterman (Ed.), *Readings in Translation Theory*, Finland, Oy Finn Lectura Ab, pp. 99-104.

KOLLER, W. (1995): "The Concept of Equivalence and the Object of Translation Studies", *Target*, 7 (2), pp. 191-222.

KRINGS, H. (1987): "The Use of Introspective Data in Translation", *Introspection in Second Language Learning*, Færch and Kasper (Eds), Clevedon/Philadelphia, Multilingual Matters, pp. 159-176.

KUHN, T. (1970): *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.

LAKOFF, G. (1987): *Women, Fire and Dangerous Things*, Chicago, University of Chicago Press.

LAVIOSA, S. (1998): "The English Comparable Corpus: a Resource and a Methodology", Bowker, Cronin, Kenny and Pearson (Eds.), *Unity in Diversity? Current Trends in Translation Studies*, Manchester, St. Jerome Publishing.

LAVIOSA-BRAITHWAITE, S. (1996): "Comparable Corpora: Towards a Corpus Linguistic Methodology for the Empirical Study of Translation", Thelen and Lewandowska-Tomaszczyk (Eds), *Translation and Meaning. Part 3*, Proceedings of the Maastricht Session of the 2nd International Maastricht- Lódz Duo Colloquium on "Translation and Meaning", Maastricht, The Netherlands, 19-22 April 1995, Maastricht, Hogeschool Maastricht School of Translation and Interpreting, pp. 153-163.

LAVIOSA-BRAITHWAITE, S. (1997): "Investigating Simplification in an English Comparable Corpus of Newspaper Articles", Klaudy and Kohn (Eds), *Transferre Necesse Est*, Proceedings of the 2nd International Conference on Current Trends in Studies of Translation and Interpreting 5-7 September, 1996, Budapest, Hungary, Scholastica, pp. 531-540.

LEECH, G. (1992): "Corpora and Theories of Linguistic Performance", Svartvik (Ed.), *Directions in Corpus Linguistics,* The Hague, Mouton de Gruyter, pp. 105-122.

MALMKJÆR, K. (1993): "Who Can Make *Nice* a Better Word than *Pretty*?: Collocation, Translation, and Psycholinguistics", Baker, Francis and Tognini-Bonelli (Eds), *Text and Technology: In honor of John Sinclair*, Amsterdam/Philadelphia, John Benjamins, pp. 213-232.

MARMARIDOU, A.S.S. (1996): "Directionality in Translation Processes and Practices", *Target*, 8 (1), pp. 49-73.

MCENERY, T. and A. WILSON (1996): *Corpus Linguistics*, Edinburgh, Edinburgh University Press.

NORD, C. (1991): *Text Analysis in Translation. Theory, Methodology and Didactic Application of a Model for Translation-oriented Text Analysis*, Amsterdam/Atlanta, Rodopi.

ØVERÅS, L. (1995): *In Search of the Third Code. An Investigation of Norms in Literary Translation*, Unpublished Thesis, University of Oslo.

PYM, A. (1995): "European Translation Studies. 'Une science qui dérange,' and Why Equivalence Needn't be a Dirty Word, *TTR*, III (1), pp. 153-176.

QUIRK, R. (1992): "Postscript. On Corpus Principles and Design", Jan Svartvik (Ed.), *Directions in Corpus Linguistics,* The Hague, Mouton de Gruyter, pp. 457-469.

REISS, K. (1969): "Textbestimmung und Übersetzungsmethode. Entwurf einer Texttypologie", *Ruperto Carola*, 21 (46), pp. 69-75.

REISS, K. (1976): *Texttype und Übersetzungsmethode. Der opertive Text*, Kronberg/Ts. Scriptor, Heidelberg, Groos.

REISS, K. (1981): "Type, Kind and Individuality of Text. Decision-making in Translation", *Translation Theory and Intercultural Relations* special issue of *Poetics Today*, 2 (4), Even-Zohar and Toury (Eds), pp. 121-131.

REISS, K. (1989): "Text Types, Translation Types and Translation Assessment", Chesterman (Ed.), *Readings in Translation Theory*, Finland, Oy Finn Lectura Ab, pp. 105-115.

REISS, K. and Hans VERMEER (1984): *Grundlegung einer allgemeinen Translationstheorie*, Tübingen, Narr.

ROSCH, E. (1973): "Natural Categories", *Cognitive Psychology*, 4, pp. 328-350.

ROSCH, E. (1975a): "Cognitive Reference Points", *Cognitive Psychology*, 7, pp. 532-547.

ROSCH, E. (1975b): "Cognitive Representations of Semantic Categories", *Journal of Experimental Psychology: General*, 104, pp. 192-233.

ROSCH, E. (1977): "Human Categorization", *Studies in Cross-Cultural Psychology*, London, Academic.

ROSCH, E. (1978): "Principles of Categorization", *Cognition and Categorization,* Rosch and Lloyd (Eds), Hillsdale, N.J., Lawrence Erlbaum Associates, pp. 27-48.

ROSCH, E. (1981): "Prototype Classification and Logical Classification: The Two Systems", Scholnick (Ed.), *New Trends in Cognitive Representation: Challenges to Piaget's Theory*, Hillsdale, N.J., Lawrence Erlbaum Associates, pp. 73-86.

ROSCH, E., MERVIS, C., GRAY, W., JOHNSON, D. and P. BOYES-BRAEM (1976): "Basic Objects in Natural Categories", *Cognitive Psychology*, 8, pp. 382-439.

SÉGUINOT, C. (1991): "A Study of Student Translation Strategies", Tirkkonen-Condit (Ed.), *Empirical Research in Translation and Intercultural Studies*, Selected papers of the TRANSIF seminar, Savonlinna, 1988, Tübingen, Gunter Narr, pp. 79-88.

SINCLAIR, J. (1992): "The Automatic Analysis of Corpora", Svartvik (Ed.), *Directions in Corpus Linguistics,* The Hague, Mouton de Gruyter, pp. 379-397.

TIRKKONEN-CONDIT, S. (1988): "Interpretation of Meaning in Translation", Laurinen and Kauppinen (Eds), *Problems in Language Use and Comprehension*, AFinLa Yearbook 1988, Helsinki, Publications de l'Association finlandaise de linguistique appliquée, pp. 145-155.

TIRKKONEN-CONDIT, S. (1990): "Professional vs. Non-professional Translation: A Think-Aloud Protocol Study", Halliday, Gibbons and Nichols (Eds), *Learning, Keeping and Using Language: Selected Papers from the Eigth World Congress of Applied Linguistics,* Sydney, Australia 16-21 August 1987, Amsterdam, John Benjamins, pp. 381-394.

TIRKKONEN-CONDIT, S. (1992): "The Interaction of World Knowledge and Linguistic Knowledge in the Processes of Translation. A Think-Aloud Protocol Study", *Translation and Meaning. Part 2*, Thelen and Lewandowska-Tomaszczyk (Eds), Maastricht, Rijkshogeschool Maastricht, Faculty of Translation and Interpreting, pp. 433-440.

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

TOURY, G. (1984): "The Notion of 'Native Translator' and 'Translation Teaching'", Wilss and Thome (Eds), *Die Theorie des Übersetzens und ihr Aufschlussert für die Übersetzungs- und Dolmetschdidaktik*, Tübingen, Gunter Narr, pp. 186-195.

TOURY, G. (1985): "A Rationale for Descriptive Translation Studies", Hermans (Ed.), *The Manipulation of Literature: Studies in Literary Translation*, London & Sydney, Croom Helm, pp. 16-41.

TOURY, G. (1995): *Descriptive Translation Studies and Beyond*, Amsterdam/Philadelphia, John Benjamins.

VERMEER, H. (1989): "Scopos and Commission in Translational Action", *Readings in Translation Theory,* Chesterman (Ed.), Finland, Oy Finn Lectura Ab, pp. 173-187.

WILSS, W. (1982): *The Science of Translation: Problems and Methods*, Tübingen, Narr.

WITTGENSTEIN, L. (1958): *Philosophical Investigations*, Oxford, Blackwell.