Translation Universals

Universals

Do they exist?

EDITED BY

Anna Mauranen
Pekka Kujamäki

Translation Universals

Benjamins Translation Library

The Benjamins Translation Library aims to stimulate research and training in translation and interpreting studies. The Library provides a forum for a variety of approaches (which may sometimes be conflicting) in a socio-cultural, historical, theoretical, applied and pedagogical context. The Library includes scholarly works, reference works, post-graduate text books and readers in the English language.

Volume 48

Translation Universals: Do they exist?
Edited by Anna Mauranen and Pekka Kujamäki

Translation Universals

Do they exist?

Edited by

Anna Mauranen
University of Tampere

Pekka Kujamäki
University of Joensuu

Table of contents

Introduction

The search for universals of translation has experienced a surge of research interest since the mid-nineties, in particular since the advent of electronic corpora as research tools in translation studies. The seminal paper was Mona Baker's (1993) article where she suggested that large electronic corpora might be the ideal tool for investigating the linguistic nature of translations: either in contrast to their source texts or in contrast to untranslated target language texts. Baker saw in electronic corpora a useful testbed for a series of hypotheses on universal features of translation that had been put forward by other scholars on the basis of small-scale, manually conducted contrastive studies only. Included in her list were features such as a tendency towards explicitation (Blum-Kulka 1986; Toury 1991a), disambiguation and simplification (Blum-Kulka & Levenston 1983; Vanderauwera 1985), growing grammatical conventionality and a tendency to overrepresent typical features of the target language (Toury 1980; Vanderauwera 1985; Shlesinger 1991) as well as the feature of cleaning away repetitions from translations (Shlesinger 1991; Toury 1991b). Since this article, the idea of linguistic translation universals has found a place at the centre of discussion in translation studies.

The idea of translation studies searching for general laws and regularities is not new; the best-known advocate for general laws of translation has been Gideon Toury (1980, 1995), who proposed this as a fundamental task of descriptive translation studies. Similarly, more recently Andrew Chesterman (e.g. 1998, 2000) has wished to see translation studies as a rigorously scientific pursuit, seeking generalisations like any other science. A clearly linguistic flavour to the issue has been added by those who have suggested that translated language is a kind of 'hybrid language' (see e.g. Trosborg 1996 and 1997; Schäffner & Adab 2001), or a 'third code' (Frawley 1984).

The issue remains highly controversial: while some scholars (e.g. Laviosa-Braithwaite 1996) claim that they have found clear support for hypotheses concerning general linguistic properties of translated language such as simplification, others (e.g. Tymoczko 1998; Paloposki 2002) maintain that the very idea of making claims about universals in translation is inconceivable since we

have no way of capturing translations from all times and all languages. Others, again, are proposing new subtypes of universals (Chesterman 2001), questioning or further developing already established concepts, (e.g. Toury 2001, Klaudy 2001) or wondering if the term was felicitous after all (Baker 2001). The discussion is very much alive, and to fuel it further, we are now rapidly accumulating evidence from actual data which demands interpretation.

In linguistics, universals have been discussed for quite a while, and it has become clear that a fruitful study of language universals needs to take into account several different kinds, including important tendencies shared by many languages, not only 'absolute' universals, or, as Greenberg et al. (1966) put it in their classic 'Memorandum concerning language universals': "Language universals are by their very nature summary statements about characteristics or tendencies shared by all human speakers." Such an extended view – which includes tendencies – also seems to suit translation studies. Moreover, distinctions between universals which can be traced back to general cognitive capacities in humans, and those which relate linguistic structures and the functional uses of languages (see, Comrie 2003) provide food for thought for the study of translations and characteristics of translated language as well. We may want to differentiate our search for that which is most general first of all in cognitive translation processes, secondly, the social and historical determinants of translation, and finally, the typical linguistic features of translations. However, the greatest part of empirical investigation into translation universals has so far focused on linguistic characteristics – while theoretical discussion has concerned the plausibility, kinds and possible determinants of universal tendencies. There is a need to clarify the issues and also to bring together these angles, to the extent that it is possible.

Clearly, the quest for translation universals is meaningful only if the data and methods we employ are adequate for the purpose. The value of universals in deepening our understanding of translation lies in developing theory and accumulating evidence from all the three main domains that are relevant to universals: cognitive, social, and linguistic. There is therefore no reason to subscribe to any methodological monism, even though the impetus for systematic linguistic research of translation universals originated in corpus studies. There are good reasons to expect corpus methods to make an important contribution to the field in that they allow comparisons of linguistic features on a large scale; this goes both for the more traditional approach of comparing translations with their source texts (parallel corpora) and the more recent discovery of the potential in comparing translations to similar texts written originally in the target language (comparable corpora). One of

the main methodological principles in an ambitious domain like this is to keep in mind the diversity of languages, and not draw excessively hasty conclusions on the basis of comparing typologically very close languages only, or a very small range of languages.

The present volume is a selection of articles from an international conference with the same topic as the book, "Translation Universals – Do They Exist?" held in Savonlinna, October 2001, on questions relating to translation universals. Despite the uniform focus on the topic, it comprises a number of different approaches from theoretical discussion of the issues to empirical studies testing some of the main hypotheses put forth so far. The research field is still very new, as empirical work only seriously began in the late nineties. Several papers discuss the established hypotheses on universals in the light of recent work in different languages, and some move on to test new hypotheses that have emerged out of the research carried out in the last two or three years. One of the central issues is the role of interference in relation to translation universals, and a number of suggestions are made as to its position, based on various empirical approaches. Most studies report work based on large translational corpora, which have begun to appear in many languages now, with applications to translator education also included. The papers cover a number of source and target languages, which makes a welcome change in the heavily English-dominated field.

The volume is divided into four main sections, according to the main foci of the papers. Those in the first section, *Conceptualising Universals*, address issues concerning the notion of universals and universality, and the extent to which this is appropriate or fruitful as an avenue for translation studies to take. The first two articles, by Gideon Toury and Andrew Chesterman, discuss the concept of universals, reflecting upon the possibility, and indeed desirability, of discovering them in translations. Both stress the demanding nature of the enterprise, and the methodological difficulties involved. Nevertheless, both also see the search for universals as an important step forward for translation studies, particularly as regards the character and credibility of translation studies as a 'science'. Moreover, both welcome corpus-based work as a major road towards progress in the field, while neither is actively personally involved in corpus-based studies. Gideon Toury's opening article discusses the roles of different levels of abstraction in discovering regularity, and posits probabilistic statements at the highest level of generality. He then raises the question whether probabilistic propositions, or conditioned regularities, are the best we can hope for in descriptive translation studies, and if this is so, are these the universals we have been looking for. The value of the concept of universals for Toury lies not

in the possible existence of such laws, but in their explanatory power, which, at least for the time being, shows great promise. Toury prefers the term 'laws' to 'universals', but concedes to talk about universals in the present context, without too much concern.

Andrew Chesterman continues the thread of thought stemming from the quest for generalities which characterises all science. He considers the different ways in which translation studies have sought the general, distinguishing what he calls the prescriptive route, the pejorative route, and the descriptive route. The contributions and problems of each are discussed, with the main focus on the currently prevailing descriptive study of translation, where further distinctions are made, such as the very useful one between universals which relate to the process from the source to the target text (what he calls S-universals), and those which compare translations to other target language texts (T-universals). He raises many other fundamental questions, relating to the nature of evidence, the concept of tendency, and the problem of testing very high-level hypotheses, questioning for each the current conceptualisation and terminology, which, not surprisingly, tend to vary widely and often suffer from vagueness. Finally, Chesterman invites us to go beyond descriptions, to explanations, and consider questions of causes as well as effects. He calls for wider testing of hypotheses, standardisation and operationalisation of concepts, and generation of new hypotheses.

The final paper in this section, by Silvia Bernardini and Federico Zanettin, assesses the appropriateness of corpus-based approaches to the search for universals. Terminologically, they share Toury's preference for 'law' over 'universal', although for different reasons (its better fit into the framework of Firthian linguistics). They address the issue of corpus design in view of the claims that have been made for their ability to offer a testbed for translational hypotheses at the highest levels of generalisation. The discussion is filtered through an illustrative case, the compilation of an English-Italian translational corpus, which is of the parallel corpus type, and bidirectional. The organising concept is Toury's "preliminary norms", that is, the translation policies which largely determine things like the selection of texts for translation. A survey of texts that are available in translation quickly reveals that a considerable asymmetry prevails between languages as regards the proportions of genres. Sheer overall numbers show that for a given language pair, more gets translated in one direction than the other. In addition, translations in one direction are likely to be differently biased for prestige, date of original, and other social determinants. The dilemma that follows is that comparability of the texts conflicts with the objective of reflecting the prevailing preliminary norms, although an ambi-

tious corpus would wish to incorporate both criteria. The suggested solution is a broad-based methodological approach to translational corpus compilation, which gives due recognition to the social contexts that translations reside in.

The second section, *Large-Scale Tendencies in Translated Language*, is methodologically fairly uniform in that each paper reports a corpus-based study, and addresses questions of capturing universals with this approach. Moreover, they all make use of the same corpus, the Corpus of Translated Finnish, a comparable corpus which consists of 10 million words altogether, consisting of both translations into Finnish in several genres, and comparable texts originally written in Finnish. The texts are contemporary, and include translations from a number of different source languages. The corpus, which is one of the largest comparable corpora in existence, was compiled at the Savonlinna School of Translation Studies in the last five years of the 1990s. The first of the papers, by Anna Mauranen, who was the initiator and director of the Savonlinna project, gives an account of the structure and origins of the corpus. Her paper sets out by considering the problem of interference in translation, which has been used rather carelessly and given diverse interpretations, and then moves on to explaining and trying out a procedure for comparing different corpora in search for evidence on the role of interference and transfer. A corpus comparison on an overall basis is problematic; the present solution is based on lexis and rank order, and it obviously needs other types of evidence to support or refute the findings. Nevertheless, the method yields results which suggest that translations are more similar to one another than to originals in the target language, but that translations from particular source languages and cultures differ from each other in their distance from the target language texts. This suggests that interference is a fundamental property of translations, but that not all linguistic features specific to translations are reducible to interference – other sources are required to explain the rest of the distance between translations and non-translations on the one hand, and the proximity of translations to one another.

The topic of interference is followed on by Sari Eskola, who advocates a new reading to the concept of interference as a neutral, non-pejorative term. Her theoretical interest is also in clarifying the concepts of 'norm' and 'universal' with respect to regularities, and she suggests the common term for observed regularities should be Toury's 'law', with a distinction being made between local and global laws, the latter representing universals. She has investigated the syntax of texts translated into Finnish in comparison with originally Finnish texts, with Russian and English as source languages, both typologically very distant from Finnish. Her particular focus is on non-finite constructions,

which, on the face of it, could be assumed to be typical of translations in that they offer convenient ways of overcoming syntactic differences in the source and target languages. Her findings indicate that translations, compared with original TL texts, overrepresented those SL features which had straightforward translation equivalents in the TL, but, conversely, underrepresented features which were specific to the TL. This supports Tirkkonen-Condit's (2000, this volume) hypothesis on the relative underrepresentation of unique items, and also Mauranen's (2000) findings on word combinations. Since the latter studies were based on lexis, Eskola's syntactic results provide an important support. Eskola's finding that the differences between translations from Russian and Finnish originals are greater than between translations from English vis-a-vis Finnish originals are in line with Mauranen's lexical results (this volume).

Jarmo Harri Jantunen takes up the methodological issues involved in the quest for universals with the help of comparable target-language corpora. His study is also based on a subsection of the Corpus of Translated Finnish (CTF). His particular focus is on lexical patterning, more specifically near-synonymous frequent intensifiers, but the main objective of the paper is to present a quantitative methodological solution for investigating the influence of the SL on translations. The three-phase method of comparisons is enabled by the compilation principles of the CTF, and Jantunen takes pains to explore the suitability of various statistical measures for discovering meaningful regularities in the data in a reliable way. His findings are interestingly complex in that the very small selection of near-synonyms showed different patterning both in terms of collocations and colligations, and the main conclusion is that it is imperative to continue fine-tuned research into specific cases to be able to appreciate the extension of SL influence and other determinants of difference and similarity in translated and untranslated language.

The third section, *Testing the basics,* is devoted to papers in which some basic assumptions on the specificity of translated language are tested with different parallel and comparable corpora. The section is opened by Per-Ola Nilsson, who reports on a methodologically rigorous corpus-driven study of translation-specific lexicogrammar in texts translated from English into Swedish. The quantitative comparison of original and translated Swedish reveals that in the translated text corpus, the grammatical word *av* as well as many collocational patterns and frameworks including *av* were significantly overrepresented. Nilsson uses the fiction part of the English-Swedish Parallel Corpus (ESPC), which with its aligned subcorpus enables him to move on to the search for causes for this overrepresentation. The analysis shows a strong structural correspondence between English sources and Swedish translations: the transfer

of several frequent SL patterns give rise to these frequency differences between translated and non-translated Swedish.

One of the assumed universals of translation is explicitation. The hypothesis is used to refer either to the process or strategies of making translations more explicit than their source texts, or to the tendency of translated texts to exhibit a higher degree of explicitness than original, non-translated texts of the same TL. To cater methodologically for both assumptions, Vilma Pápai analyses in her paper a combination of parallel and comparable corpora of Hungarian and English literary and non-literary texts (the ARRABONA corpus). First, the analysis of translators' shifts in the parallel corpus reveals a series of frequent explicitation strategies on different linguistic levels. At the second stage, these strategies are taken up for closer analysis in a comparable corpus of Hungarian. The results provide evidence in support of the above hypotheses on explicitation as a characteristic feature of the translation process and on the explicitness of translated texts as compared to non-translated ones. In contrast to Pápai's further hypothesis, however, the quantitative data does not point to any significant differences between the analysed genres, i.e. between literary and non-literary texts. Finally, Pápai investigates the lexical complexity of translations and non-translated texts (type/token ratio) and suggests a connection between various explicitation strategies (e.g. lexical repetition, addition of conjunctions, filling in ellipsis) and simplification – another alleged universal of translation.

The second paper dealing with explicitation is written by Tiina Puurtinen. In contrast to Pápai, Puurtinen concentrates only on explicitation as "a potentially distinctive quality of translations in comparison with non-translated TL texts of the same type", in this case contemporary children's literature. Potential manifestations of this quality are the explicit signals of clausal relations, which offer themselves for use in translated texts as alternatives to other rather implicit and complex realisations such as non-finite constructions (NCs). Puurtinen's earlier research on translated children's literature showed that even though NCs are likely to decrease the readability of a text as well as the facility with which it can be read aloud, and also to make the text more difficult for children to understand, they nevertheless are very common and significantly more frequently used in translated than in non-translated children's fiction. Puurtinen interprets this as evidence contrary to the hypothesis of explicitation being a universal tendency. Her basic research question is, then, whether this feature correlates with infrequent use of explicit connectives in translated children's literature. Her findings remain inconclusive, since no clear correlation was found between low connector use and high NC use. She

suggests that subtler differences obtain between different subcorpora and in the specific usage of different connectors.

Both Puurtinen and Sonja Tirkkonen-Condit, whose article closes the third section, use the comparable Corpus of Translated Finnish as their data. Tirkkonen-Condit's point of departure is the hypothesis on allegedly universal overrepresentation of those linguistic features in translation that are typical of the target language. She challenges this view by comparing the frequencies of a number of Finnish verbs of sufficiency as well as of some clitic pragmatic particles – two examples of "unique items" that are very typical of the Finnish language but lack linguistic counterparts in English, the source language here. This is why – as Tirkkonen-Condit's hypothesis reads – they do not suggest themselves as first choices for translation. The hypothesis of the relative underrepresentation of target language-specific features is therefore a new candidate among universals. The author discusses the overall results of the comparison and combines them with observations on the translation process in general.

The concept of "unique items" is taken up by Pekka Kujamäki, who opens the fourth and final section *Universals in the translation class*. To show his students the function of Toury's "law of interference" Kujamäki compares students' translated Finnish with their English and German source texts *and* with their non-translated language use as revealed by a small cloze test. The experiment indicates a strong adherence to the surface structure of the source texts in student translations, in which – neatly in compliance with Tirkkonen-Condit's above hypothesis – straightforward lexical or dictionary equivalents of the English and German stimuli suggest themselves as translations much more easily than the more natural sounding "unique items" of the target language.

Finally, Riitta Jääskeläinen closes the volume with a report on a research project in progress which aims at discovering whether and in which ways students of translation can be made aware of the stylistic function of repetition in texts. Her point of departure is an observation in the translation class which complies with one assumed translation universal, namely, that students tend to clean away repetition from their translations. Jääskeläinen compares students' translations that are produced with or without "sensitivity training", and relates their strategies to different mechanisms at work in translation.

A recurrent issue in many if not all of the papers in this volume is whether the term 'translation universal' is felicitous, and many writers seem to be somewhat uneasy about it, suggesting other, related terms according to personal preferences. However, they do not object seriously enough to deny the usefulness of the concept as a tool, at least provisionally, at least for the present.

For many, the term universal is perhaps too radical, too abrupt, too absolute. Such objections may result in a general preference for another term, such as 'regularity', 'law', or 'tendency', depending on how far we dare to tread. We may ask for example to what extent the postulation of universals is restraining our focus on mainstream, prototypical translation in contemporary developed world to the exclusion on more marginal and historical translation practices.

Given that the accumulated evidence is still scarce, it is impossible to tell how general we can get in our descriptions – without ending up with truisms such as 'all translations involve two linguistic codes' or other general statements which follow from the definition of translation. Disputes about such uninformative top-level generalisations would then boil down to controversies about definitions of translations. Clearly, our theoretical framework largely determines our possibilities of seeing the object, thus we cannot naively wait for the evidence to accumulate until there is enough to resolve the issues. Yet by making strong claims in the field, and by imposing strong frameworks on our data, we stand a chance of seeing the limits of a new approach, as well as its strengths. We hope that this volume makes a contribution to the search for generalities in translation studies, the methodological solutions available, and the emerging evidence on the kinds of generalities that research on a larger scale than before is bringing forth, enabling us to fine-tune, modify, and question earlier hypotheses. On a more practical but no less important level, the applicability of the hypotheses and findings to translator education is always a concern for translation studies.

References

Baker, Mona (1993). Corpus linguistics and translation studies – Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology. In Honour of John Sinclair* (pp. 233–250). Amsterdam: John Benjamins.

Baker, Mona (2001). Patterns of idiomaticity in translated vs. original English. Paper given at the Third EST Congress Translation Studies: Claims, Changes and Challenges, August 30 – Sept.1, 2001, Copenhagen.

Blum-Kulka, Shoshana (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17–35). Tübingen: Gunter Narr.

Blum-Kulka, Shoshana, & Levenston, Eddie A. (1983). Universals of lexical simplification. In C. Faerch & G. Kasper (Eds.), *Strategies in Interlanguage Communication* (pp. 119–139). London/New York: Longman.

Chesterman, Andrew (1998). Causes, translations, effects. *Target, 10* (2), 201–230.

Chesterman, Andrew (2000). A causal model for Translation Studies. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects* (pp. 15–27). Manchester: St. Jerome Publishing.

Chesterman, Andrew (2001). Classifying translation universals. Paper given at the Third EST Congress Translation Studies: Claims, Changes and Challenges, August 30 – Sept.1, 2001, Copenhagen.

Comrie, Bernard (2003). On explaining language universals. In M. Tomasello (Ed.), *The New Psychology of Language*. Vol. 2 (pp. 195–209). Mahwah, NJ: Lawrence Erlbaum.

Frawley, William (1984). Prolegomenon to a theory of translation. In W. Frawley (Ed.), *Translation. Literary, Linguistic and Philosophical Perspectives* (pp. 159–175). Newark: University of Delaware Press.

Greenberg, Joseph H. (1966). *Universals of Language*. 2nd edition. Cambridge, Mass. The MIT Press.

Klaudy, Kinga (2001). The asymmetry hypothesis. Testing the asymmetric relationship between explicitations and implicitations. Paper given at the Third EST Congress Translation Studies: Claims, Changes and Challenges, August 30 – Sept.1, 2001, Copenhagen.

Laviosa-Braithwaite, Sara (1996). *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*. Unpublished PhD thesis. UMIST, Department of Language Engineering.

Mauranen, Anna (2000). Strange Strings in Translated Language: A Study on Corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects* (pp. 119–141). Manchester: St. Jerome Publishing.

Mauranen, Anna (this volume). Corpora, Universals and Interference. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Paloposki, Outi (2002). *Variation in Translation. Literary Translation into Finnish 1809–1850*. Unpublished PhD thesis, University of Helsinki.

Schäffner, Christina, & Adab, Beverly (2001). The idea of the hybrid text in translation: Contact as conflict. *Across Languages and Cultures.* Special issue on Hybrid Texts and Translation. C. Schäffner & B. Adab (guest Eds.), 167–180.

Shlesinger, Miriam (1991). Interpreter latitude vs. due process. Simultaneous and consecutive interpretation in multilingual trials. In S. Tirkkonen-Condit (Ed.), *Empirical Research in Translation and Intercultural Studies* (pp. 147–155). Tübingen: Gunter Narr.

Tirkkonen-Condit, Sonja (2000). In search of translation universals: Non-equivalence or 'unique' items in a corpus text. Paper presented at the conference Research Models in Translation Studies, UMIST, 28–30.4.2000, Manchester.

Tirkkonen-Condit, Sonja (this volume). Unique Items – Over- or Under-Represented in Translated Language? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

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

Toury, Gideon (1991a). Experimentation in translation studies: achievements, prospects and some pitfalls. In S. Tirkkonen-Condit (Ed.), *Empirical Research in Translation and Intercultural Studies* (pp. 45–66). Tübingen: Gunter Narr.

Toury, Gideon (1991b). What are descriptive translation studies into translation likely to yield apart from isolated descriptions. In K. M. van Leuven-Zwart & T. Naaijkens (Eds.), *Translation Studies: The State of the Art* (pp. 179–192). Amsterdam: Rodopi.

Toury, Gideon (1995). *Descriptive Translation Studies – and Beyond*. Amsterdam: John Benjamins.

Toury, Gideon (2001). Probabilistic Explanations in Translation Studies: Universals – Or a Challenge to the Very Concept? Paper given at the Third EST Congress Translation Studies: Claims, Changes and Challenges August 30 – Sept. 1, 2001, Copenhagen.

Trosborg, Anna (1996). Hybrid texts as a product of a supranational discourse. Paper given at AILA 96, the 11th International Congress of Applied Linguistics, 4–9.8.1996, Jyväskylä, Finland.

Trosborg, Anna (1997). Translating Hybrid Political Texts. In A. Trosborg (Ed.), *Analysing Professional Genres* (pp. 145–158). Amsterdam: John Benjamins.

Tymoczko, Maria (1998). Computerized corpora and Translation Studies. *Meta, 43*(4), 652–659.

Vanderauwera, Ria (1985). *Dutch Novels Translated into English: The Transformation of a Minority Literature*. Amsterdam: Rodopi.

Conceptualising universals

Probabilistic explanations
in translation studies

Welcome as they are, would they qualify
as universals?*

Gideon Toury
Tel Aviv University

> Part of the meaning of a [...] system is the relative probability of its terms.
>
> (Halliday 1991a: 48)

There is no doubt a vast array of factors which have the capacity to influence the selection of a particular translational behavior or its avoidance. Although we have no real list, it is clear that this array is *heterogeneous* in its very nature: some of the variables are cognitive, others cross-linguistic or socio-cultural, and there are no doubt more. Due to this vastness and heterogeneity, there can be no *deterministic* explanation in Translation Studies. First of all, there seem to be no single factor which cannot be enhanced, mitigated, maybe even offset by the presence of another. Secondly, the different variables are present (and active) all at once rather than one by one, so that there are always several factors interacting, and hence influencing each other as well as the selected behavior. In an attempt to escape the trap of deterministic reasoning I suggested a different format of explanation; namely, a *conditioned*, and hence *probabilistic* one, and defined the ultimate aim of TS as moving gradually, and in a controlled way, towards an empirically-justified theory which would consist in a system of interconnected, even interdependent probabilistic statements. The present paper will return to all these issues with the intention of asking whether, welcome as they certainly are, such explanations qualify as "universals of translational behavior" and, if not, whether there are any other candidates for universal-ship.

1.   Introduction

Even though a deliberate search for regularities has long been recognized as an inherent feature of the endeavor of science, the quest for universals is anything but common practice among translation scholars. In fact, it is almost the other way around: there have been, there are, and there will probably always be many who would value differences over similarities any time. Some would even declare not mere lack of interest in, but even hostility towards the very idea of searching for recurrent patterns, purporting as they do to show what is *unique* to whatever they set their heart on, at a particular moment.

It is not difficult to sympathize with them either. After all, we have all been in their shoes once. At the same time, I cannot but wonder how those who subscribe to such a position think they are ever going to know what is truly unique (and I do not doubt it that, although some instances of translation are certainly less unique than others, there *is* a measure of uniqueness in all of them) unless they have at least some idea of what their immediate object of study *shares* with other possible objects. Or is there anyone who would still maintain that translation is *erratic* in its nature, so that shared features, if and when encountered, represent a mere accident? – Because, sooner or later, shared features, at one level or another, are bound to emerge.

True, the first cases one studies often seem fraught with revelations. At times, almost everything may look like a genuine discovery. However, this is just an optical aberration, the reflection of a beginner's lack of previous experience, not to say naïveté. Thus, as one increases one's knowledge, or expands the field one takes into account, certain phenomena start repeating themselves and gradually become more predictable than others. Any further expansion of the object of study, especially if it is done systematically (i.e. on the basis of an explicit criterion, or set of criteria, which also lend themselves to control), would contribute towards undermining the (evidently erroneous) first impression of uniqueness, until it is finally reversed. Unfortunately, by that time, many would have stopped doing active research in translation or left academia altogether, surrendering the field to (inevitably naïve) newcomers. The latter would go through the same initiation process again, albeit (probably) at a somewhat quicker pace, due to some permanent impressions left in the field by previous generations of scholars. Those few who would stay with us, on the other hand, will no longer experience too many surprises. For them, almost everything, certainly everything *of essence*, will have become highly predictable.

Be the balance between the two positions among translation scholars as it may, I wish to proceed from a naïve assumption myself; namely, that all

those who – of their own free will – chose to attend a Workshop dealing with "Translation Universals" (or read its Proceedings) share at least some basic willingness not only to accept the existence of *regularities* in translational behavior and the idea of searching for them, but also to give the notion of '*universals*' a shot, or at least suspend their disbelief for a while. At the same time, it is important to bear in mind that, while universals do presuppose regularities, the reverse does not necessarily hold: It is one thing to say that certain regularities were found *in* translation, and something quite different – to claim that the observed regularities are there *because* it is translation.

Thus, the transition from 'regularities' in general to narrower, more specific 'universals' is not, nor can it be done automatically. Rather, it requires research work which will take its cue from the demands we would like to make on universals. *Quantity* will evidently play an important role in the transition, but it is not at all necessary that it would be made on quantitative grounds alone.

In what follows, I will therefore say nothing about possible justifications for the search for universals in the field of translation as such, nor would I submit any individual candidate for universal-ship to detailed scrutiny and analysis:[1] no disagreement on the status of any single proposition as a possible universal should be taken to invalidate the concept itself or render the quest for translation universals unfounded, let alone illegitimate. Finally, I will attempt no classification of possible universals either: even if I were a lover of nomenclature (which I have never been, especially if the nomenclature is established in advance, and on purely speculative grounds), the time doesn't seem ripe. My main concern will rather be with *the transition itself* from regularities to universals. In this context, I will tackle two main issues:

– the place where translation universals might be located, and
– the form such universals would be given, if and when their existence and usefulness have been established.

2.   Universals should not be sought on too concrete a level

Let me start with the obvious:

I assume it is clear (and hence agreed) that universals should not be sought on too concrete a level, where many of the identified regularities can quite easily be given an *exact numerical value* and expressed as frequency. This is mainly true of individual instances of behavior, especially the behavior of single translators in single acts vis-à-vis particular, low-level phenomena which are relatively easy to delimit and detect. Here, frequencies can sometimes – albeit

seldom – be either 0 or 1: types of behavior which, in principle, could have occurred may *never* have been opted for, in practice, while others may *always* be present, irrespective of anything.

However, not only individual instances of translational behavior, but that of definable bodies as well – whether groups of persons (e.g. so-called 'schools' of translators) or groups of acts (or their observable results, i.e. corpora of 'assumed translations')[2] – don't seem to constitute candidates for elevation to universal-ship; certainly not until the findings have been *relativized* in view of the factors defining the group in question, which would involve a considerable distancing of the vantage point, a kind of 'zoom out' motion resulting in the possibility of regarding an extended field while losing sight of the minute details. Thus, even if a low-level regularity will later be shown to be an instantiation of some higher-level universal, the instantiation itself as revealed when regarded from a short distance will always be *local*, i.e. norm-governed or idiosyncratic, typical of a group or an individual, depending on the size and/or heterogeneity of the object of study.

It is not that norms, even many of the idiosyncrasies, do not imply regularities, then, because both of them do. It is only that the regularities they imply are not *general* enough to be regarded as universals, in terms of either the population or the scope of the phenomenon examined. When a low-level phenomenon is tackled within a more extensive, and especially more heterogeneous corpus, the normal result seems to be an immediate *drop* of frequency value. This value will rise again when that phenomenon is no longer viewed in itself, but as one of a number of possible instantiations of a higher-level, more general category (e.g. a recurring replacing word vs. a speech organizer, or a metaphor).

Consider the following series of research tasks, which is both simplified and highly partial, skipping many of the possible interim links:[3]

Hebrew translational replacements of:
the speech organizer *well* (= a recurring *word*)
in one particular story translated from English
in all the stories translated from English
in one particular year
decade
generation
millennium
in English texts of other types
in an English text in general

Hebrew translational replacements of the English speech organizer *oh* (= another recurring *word*)
under the same sets of circumstances

Hebrew translational replacements of an English speech organizer (= a whole *functional category* as realized in one particular language). Its treatment presupposes the establishment of correlations between various realizations like *well* and *oh*, including combinations such as *oh, well* or *well, well.*

Hebrew translational replacements of speech organizers in general (= a *super-linguistic category*). Presupposes work on a number of different target languages first the translational treatment of semantically depleted lexical items (a category which is *more inclusive* still) in the Hebrew context changes in all the above practices over time:

in phylogenesis
in ontogenesis

the whole list repeated for a [series of] different target language[s] the whole list repeated once again for the sum-total of target languages (or for 'target language' in general, and hence maybe of *translation as such*)[4]

The main point relevant to our concerns should have become visible by now. One way of making it more explicit would be to adopt the distinction between 'regularities of performance' and 'regularities in the system': the first one would be expressed as *frequencies* (e.g. "the frequency of the occurrence of the lexeme *u-vexen* as a Hebrew replacement of English *well* in the translation of text X by translator Y is 99/100"), the other one – in *probabilistic* terms (e.g. "the likelihood that an existing Hebrew speech organizer will replace an English one in prose fiction translations of the 1950s is three times lower than it would be in the 1960s").

It is clear that the two notions are distinct, but not unconnected. In fact, "Frequency in text is the instantiation of probability in the system", as Michael A.K. Halliday put it in a seminal article on the use of probabilistic interpretations in linguistics (1991a: 42). In other words, "The system may have infinite potential; but it engenders a finite body of text, and text can be counted" (1991a: 41).[5]

Frequencies can thus be tackled in a *direct* way, on the basis of surface realizations of more abstract categories, whereas probabilities will always be a number of further steps removed. Actually, says Hans Reichenbach in the 1948 edition of his *Theory of Probability*, "probability [is] the limit of the infinite series representing the frequency", where 'limit' is used in a purely

mathematical sense. Put in slightly different terms, one could perhaps say that frequency applies first and foremost to things *past*, whereas probability makes a claim for validity in the *future*. Be that as it may, in a field like translation, the best, if not the only way to go about estimating "probabilities for terms in [...] systems" is to proceed from "observed frequencies in [a] corpus" (Halliday 1991a: 42).

3.   Universals shouldn't be sought on too high a level either

On the other hand, there are also levels of generality which seem to be *too high* for the kinds of universals we are searching for, especially if we wish those universals to add something to our knowledge and understanding of translation and to be non-trivial, at the same time. Thus – to me, at least – sweeping statements of the form TRANSLATION INVOLVES EXPLICITATION,[6] or SIMPLIFICATION, or NORMALIZATION, are at least suspicious, in that respect, be they given a 'weaker' or a 'stronger' reading. (One reading or another will always have to be applied, due to the inherent vagueness of the formulations; see Toury forthcoming.)

If such a proposition is understood as a claim for *exclusiveness* – for instance, if TRANSLATION INVOLVES EXPLICITATION is taken to imply that it is *only* instances of explicitation that will be encountered, to the exclusion of non-explicitation, let alone implicitation – then the claim is obviously *false*. In fact, it is not even the case that, in any individual instance of translation, *more* examples of explicitation than implicitation will occur.

Some will no doubt argue, at this point, that claims of this kind should not be taken to refer to 'translation' in general, but to something they would call the 'typical', maybe even 'prototypical' translation (e.g. Halverson 2000). However, what constitutes [proto]typicality in the field of translation is far from self-evident and therefore such a notion is not all that easy to work with. In fact, its elucidation, should one wish to use it, would form an integral part of the very hunt for universals rather than serving as a starting point for it.[7]

By contrast, if this proposition is understood to simply state that cases of explicitation *can* be found in translated texts – alongside cases of non-explicitation and implicitation, that is – it would simply be stating the obvious; and I would very much doubt that, by formulating it, the requirements of non-triviality and expansion of knowledge and understanding would have been fulfilled. What is even worse, this 'neutralizing' formulation can easily be taken to imply that the two opposites – explicitation and implicitation –

are on an *equal footing* vis-à-vis translation. Of course, this is true in as far as cases of both *can* indeed be found in individual translations, maybe even every single one of them. However, it is at least counter-intuitive when it comes to the general notion of 'translation', even '[proto]typical' translation, precisely because it lacks any indication of probability: Would one of the terms be more common, and its occurrence more predictable than its opposite?

Obviously, the less vague a statement of this kind, the easier it is to disprove it; and not only speculatively (as we have been doing so far), but on empirical grounds as well; that is, in the face of *factual evidence*. To quote Reichenbach again, "there is no need for a concept of probability which is not reducible to frequency notion". To be sure, one counter-evidence is enough to shake the universality of any such statement, and exceptions are not really difficult to find. In fact, the possibility of *fabricating* instances of counter-evidence at will may in itself undermine such statements' claim to universality, as any 'fabricated' (or 'simulated') translation is a kind of translation and nothing but translation,[8] and there is always a possibility that somebody has taken, or will be taking the same route when doing 'genuine', i.e. socially and culturally relevant translation.

4.   Would the presence of "shifts" constitute a universal?

Being 'general' is not an either/or matter, then. Rather, there seems to be a *graded scale of generality*. Let us climb another rung up that ladder and see what will happen.

You will have noticed that there is one key-feature that all statements of the format "translation involves X" have in common; namely, their predicates representing so-called *shifts*, such as explicitation, implicitation, simplification, complexification, etc., etc. are all kinds of translational shifts.[9] The obvious question to ask now, in the context of our attempt to locate the point where mere 'regularities' become proper 'universals', is: what would the status be of the common denominator itself, or the underlying proposition TRANSLATION INVOLVES SHIFTS.

My claim would be that we have entered the realm of *analytic statements*, maybe even that of flat *tautologies*, which may imply that we have now climbed a little too high.

Thus, unlike the lower-level, derivative realizations such as explicitation, implicitation, or simplification, there can be no question about the *truth* of TRANSLATION INVOLVES SHIFTS. However, this truth is by definition, so to

speak, as shifts are not just one of many possible features, but rather a *defining feature of translation*, thus forming an integral part of the very notion: claiming that a translation will necessarily reveal shifts is virtually like saying: "well, translation is translation!"

As a distinctive feature of translation, TRANSLATION INVOLVES SHIFTS is therefore not unlike propositions such as TRANSLATION INVOLVES MEMORY (which, I suspect, no one will offer as a candidate for universal-ship due to its self-evidence and non-specificity), or TRANSLATION INVOLVES NORMS, or even IS A NORM-GOVERNED ACTIVITY (which some may wish to present as one). In each case, however, the predicate (that is, the specific realization of the general notion of 'shift') draws from a different source, namely, the cross-linguistic, the cognitive and the socio-cultural, respectively.

It is not that there is anything inherently wrong with tautologies. Actually, many of them may be quite instructive, even helpful, in teaching contexts, for example. My point is only that their use adds nothing to our understanding of what translation involves. It would therefore be rather trivial – precisely what we wanted our universals not to be.

As it turns out, then, the question facing us is not really whether translation universals *exist* (as the sub-title of our Workshop had it), but rather whether recourse to the notion is in a position to offer us any *new insights*. That is to say, whether the foreseen gains (which nobody would deny) will outweigh the cumbersomeness which the introduction of a whole new categorical level into our crowded field necessarily involves.

I, for one, expect to see some gains first and foremost in terms of the ability of translation theory to account for *every individual phenomenon* occurring in the field (i.e. both to describe and explain it), if not to predict it as well, as becomes the nature of Translation Studies as an empirical discipline: will that ability be enhanced by saying that TRANSLATION INVOLVES SHIFTS of one kind or another, on one level or another? Isn't it something we have known all along? And does this known fact really represent a direct reflection of translation being translation, or are there any other mirrors we have overlooked?

Bottom line: notwithstanding the fact that such statements are certainly better candidates for universal-ship than anything we have had so far, I would go on looking for the point of transition from regularities in general to true universals; namely, somewhere in-between the idiosyncratic and norm-governed, on the one hand, and the self-evident, on the other.

5.   Probabilistic thinking in translation studies

My starting point will be an observation I made in my 1976 doctoral dissertation – a reservoir of half-baked ideas which may be worth returning to, from time to time. As the observation was made in Hebrew, I will quote from an English translation which is only partial:

> By virtue of their definition as shifts from one focal point, [...], all shifts **fall into dichotomous pairs** [...]. For instance, explicitation/implicitation, addition/loss of information, generalization/concretization, etc. Consequently [I said already then], it is possible to formulate the following rule: IF A SHIFT OCCURS, IT NECESSARILY OCCURS IN ONE DIRECTION – OR IN ITS COMPLETE OPPOSITE. The selection of one of the two options, which in themselves are given, to the extent that it is ordered [the selection, I mean], is governed by translational norms.   (Toury 1987: 6; bold-face added)

In those days, my work – theoretical, methodological and descriptive- explanatory alike – was geared towards translational norms as a theoretical notion and its use as a research tool, as well as the instantiation of such norms in a particular, well-defined field. What I failed to do was to follow my observation in any other direction. It was only in the 1990s that I realized it had something of substance to offer in terms of a possible interim zone where universal claims could maybe be made: non-trivial claims concerning regularities which are there *because* it is translation that we are looking at.

Like TRANSLATION INVOLVES SHIFTS, a statement such as TRANSLATION IS A NORM-GOVERNED ACTIVITY is analytic in nature. True, both had a certain air of novelty when they were first made explicit and added as issues to our scholarly agenda. However, whatever novelty they may have had, at the time, it seems to have worn off completely. Consequently, it is not the notion of norms itself that I wish to highlight here, but rather the idea that norms govern translational selections between modes of behavior which point in two diametrically opposite directions, involving *pairs of shifts of a complementary nature*. It is the medial zero point which is the exception, that is, it has very low probability, in most cases close to 0.

Had translational selections been random and their results, represented in and by the translated texts, totally skewed, there would have been very little one could do in terms of *explanation*, even if it were possible to come up with neat *descriptions* of individual cases (which I am not all that sure of either). Even clearer is it that there would be nothing one could contribute towards making *predictions*, be it even the kind of 'backward predictions' researchers

indulge in when they formulate hypotheses with respect to past instances of translational behaviour, to be tested against real-life acts which have already come to their end (or translated texts which are already there) but which haven't been subjected to study yet. This however doesn't seem to be the case. As I have said before, I believe there is hardly anyone today who would claim there is complete randomness in the selection of translation strategies and translational replacements, the more so as those who might have made such a claim were asked to suspend their disbelief for a while.

At the same time, I guess we would also agree, if only by intuition alone, that it is hardly the case that all modes of behavior, all phenomena, all resulting shifts, are equiprobable, that is, have the exact same initial chance of being selected, irrespective of anything. Rather, it seems that for [almost?] every complementary pair of possible ('positive' and 'negative') shifts, one of the terms – that which has higher probability – would be unmarked and the other one marked. But which would be which? This is a major issue for targeted research, especially of the empirical kind, relating to the different manifestations of the notion of 'shift'. (As already indicated, the 'neutral', medial phenomenon of 'no shift' is practically out of the game as it has a probability of [almost] nil.)

We have finally landed in the realm of probabilities, which is what I have been advocating for the last ten years or so. I can still remember a previous lecture of mine in the Savonlinna School of Translation Studies, back in 1993, which bore the first half of the present paper's title and which I never deemed ripe for publication. That lecture owed a lot to Halliday's above-mentioned (and quoted) article "Towards Probabilistic Interpretations" (1991a), where the notion was applied within the related framework of systemic-functional linguistics in a way which was then rather novel. (See also Halliday 1991b, 1993b.)

The basic idea underlying my attempts to apply probabilistic explanations to translations and translation practices was to make consistent efforts to tie together particular *modes of behavior* (or their observable results), on the one hand, with, on the other hand, an array of *variables*, whose capacity to enhance (or reduce) the adoption or avoidance of a particular behavior would be verified empirically, by means of both observational and experimental research. Even if we were to overlook the problems involved in the quantitative side of the transition from frequencies to probabilities, there are major qualitative difficulties inherent in that project, resulting not from the mere *vastness* of the said array, but first and foremost from its enormous *heterogeneity*, as the relevant variables will necessarily come from many different sources: some

will be cognitive, others linguistic (or cross-linguistic, or textual), others communicational, still others – socio-cultural; and there may well be further sources of possible variables.

It stands to reason that probabilistic reasoning and deterministic propositions would not concur: the one normally doesn't tolerate the other. In fact, there is probably no single variable affecting translation which cannot be enhanced, mitigated, maybe even offset by the presence of another. This problem is compounded by the fact that, in actual reality, there will always be more than just *two* variables influencing each other and translation behavior as a whole. After all, a translator is male or female, older or younger, more or less experienced, more or less tired, under greater or lesser time-pressure, translating into a strong(er) or weak(er) language of his or hers, well- or less-well paid, belonging to a more or less tolerant society, and so on and so forth, *all at once*, not one at a time; a tangled knot which will have to somehow be unraveled, at least for methodological purposes, and its different constituents put in some *hierarchical* order: more and less potent, more and less translation-specific, and the like.

Rather than being deterministic (i.e., having a format such as "if a then b"), the format most befitting our kind of probabilistic thinking is *conditioned* (see next Section). In principle, a reasonable ultimate goal for Translation Studies could well be to construct a system of interconnected, mutually conditioning statements, but it is certainly premature to say what such a system might look like. At this point, we don't have so much as an exhaustive *list* of possible variables, not even a speculative, untested one. We cannot even be sure that all relevant variables have already been discovered. Even less can we say with any amount of certainty what variables are stronger and weaker vis-à-vis translational behavior (in themselves, so to speak), how the members of different pairs of variables, especially those coming from different sources, act upon each other and what the results of that interaction are, or how one would move on from *pairs* to a more realistic *network* of variables and its influence on translational behavior.

6.   The format of a conditioned statement in translation studies

Since nothing can be accounted for unless we have a language for it, it would not be odd if we asked again what format a conditioned statement in Translation Studies is likely to have.

The most basic format seems to be as follows:

> If 1 and 2, and 3, and . . . ∞, then there is great likelihood that X (or else: small likelihood that no-X)

where the numbers (1, 2, 3, . . . ∞) stand for the different variables which may have an effect on the selection of a translational behavior and X – for the kind of behavior actually opted for, or, more appropriately (from the point of view of most research paradigms, which are retrospective in nature), the external manifestations of its execution, as behavior is not really observable in any direct way.

Another variant, which might be easier to use, would be:

> The presence of 1, 2, 3, . . . ∞ enhances the likelihood that X (or: reduces the likelihood that no-X)

For example,

> The coincidence of lack of experience (variable 1) and fatigue (variable 2) increases the likelihood that translational procedures will be applied to small and/or low-level textual-linguistic entities (or: reduces the likelihood that they will be applied to long and/or high-level ones, not to mention the text 'as a whole', which is a misleading concept anyway.)

(Note that no claim to validity was made. This example was intended as an illustration of the *format* only, and questions of validity seem premature anyway. The same holds for the magnitude of the said increase (that is, the probability of the occurrence of each kind of behavior under each condition), and hence its statistical significance. All these, and much more, still await targeted research.)

To be sure, even the second formulation is not really appropriate, if only because it reflects linear reasoning: the variables are taken up one by one and ordered *consecutively*, as if each one of them were operating with complete independence from all other variables. To be more acceptable, the formulation would have to take into account the above-mentioned possibility, if not likelihood, that the different variables may also affect each other. For instance:

> If 1 and 2, then the likelihood that X is greater than if only 1, and it is even greater when 3 is present too. The effect of 3 may be so strong that it completely overrides 1.

The beginning of a more elaborate version of the previous example may look like this:

> If a translator is both inexperienced (variable 1) and tired (variable 2), the likelihood that translational processing will be applied to small and/or low-level textual-linguistic entities is rather great, and it is greater still if the target culture regards the results of such behavior with considerable tolerance (variable 3). The effect of that tolerance may be so strong that *experienced* translators (variable 1 in a reversed form) would still stick to this strategy, which may therefore appear as more 'basic' to (or 'prototypical' of) translation.

From such a formulation, were it found to be valid (and I am still not saying it is!), it might prove possible to deduce a potentially high regulative capacity of cultural tolerance of textual-linguistic deviance from 'normality' in observable products of translation activities, which – when realized in actual behavior – may override many (all?) of the non-cultural variables. Intuitively, this seems to make a lot of sense (see what I said about the special status of so-called 'formal relationships' with respect to translation [e.g. Toury 1980:48]), which may render the probabilistic-conditioned apparatus as such methodologically sound, even though a lot more refinement is certainly required, and not only from the quantitative point of view. It may also shed light on the complex issue of '[proto]typical' translation.

Thus, one obvious advantage of probabilistic, conditioned formulations on the interim level we are at now is precisely that they allow systematic accounts – including elaborate explanations – of many different phenomena and groups thereof, and accounts which show a great deal of consistency, at that. They also make possible the kind of 'backward prediction' I mentioned earlier, providing that the relevant variables will indeed have been identified, weighed against each other, and brought to bear on the study. No less important, should an actual behavior emerge as different from the one predicted, it would be possible to account for the apparent deviation with no need to discard the methodological framework that yielded the frustrated expectation. My guess (and for the time being it is no more than an educated guess) is that the normal procedure will involve adding variables to the list which weren't there before, and/or refining the distinctions between different realizations of variables that were; either way, a mere *modification* rather than a complete change, which is a sign of (to me: welcome) stability.

Another crucial question concerns the *operability* of the probabilistic method. Above all, there is the question of how the probabilistic formulations will be arrived at. How will the variables be unearthed which may influence translational behavior, and how will their relative relevance and potence be

determined, alongside the ways they act upon each other? It seems safe to assume that some combination of 'top-down' and 'bottom-up' operations will be required (see e.g. Steiner 2001). The [interim] results of initial theoretical speculations will thus be examined against instances of real-world behavior and, conversely, empirical studies will be conducted, moving gradually, and in as controlled a way as possible, from individual instantiations to the culture-specific, to more and more general regularities on higher and higher levels, to generate new, or modified theoretical statements. This would no doubt enhance the pivotal role of the descriptive-explanatory branch in the evolution of Translation Studies as a whole, as foreseen in my 1995 book. It will be of special significance in the transition from a basic theory of initial possibilities to a more and more elaborate theory of 'likelihoods' (Toury 1995: 14–17). For one thing is certain: whatever we say, and whatever we do, 'hunting for regularities' is, and will always be, the name of the game.

Which brings us back to our starting point.

7.   Drawing some conclusions

Having thus brought to a close the circle I have been drawing throughout the paper, it seems a good point to stop. Let me just recapitulate – and pay the one debt I still have; namely, an explicit, if only brief consideration of the question posed in the subtitle of the paper.

Here is my summary:

1.   Regularities can be found on every level, from the individual act of translation (or translated text), up the ladder leading to the overall notion of translation, which should be applicable to all existing and possible forms of translational behavior. It is therefore not only justified, but also beneficial, to look for regularities, trying to understand not only what translation *may* involve (in general), or *does* involve (in any particular case), but also what it is, more or less, *likely* to involve, under different sets of conditions.
2.   The closer we are to the legs of the ladder, the easier it is not only to establish regularities as such, but also to quantify them – and assign at least some significance to the frequency value itself. At the extreme bottom, even 0 or 1 frequency of very low-level phenomena may sometimes be encountered, i.e. complete absence or systematic occurrence, which would no longer be the case higher up.

3. The higher we climb the ladder (that is, the bigger and/or more heteroge-
    neous our corpus is), the lower the frequency values can be expected to be
    and the lower the significance of the figures themselves – unless the issue
    under study is also extended or generalized. Statements of frequency would
    little by little be replaced by probabilistic accounts.
4. Finally, towards the top of the ladder, only probabilistic, conditioned
    propositions of a growing qualitative nature can be made, with very little
    use for numerical values, and at the very top – general statements which are
    no more than explicitations of features which are implied by the notion of
    translation itself. This is also to say that the (in)famous question of "how
    much regularity would be regarded as 'regularity'" would lose its sting.

The question now is as follows: Welcome as probabilistic and conditioned
reasoning certainly is in the context of Descriptive Translation Studies, would
the probabilistic explanations qualify as the universals we set out to find? Put
differently: if indeed ALL REGULARITIES IN TRANSLATION ARE CONDITIONED,
AND ONLY MORE OR LESS PROBABLE,[10] does it follow that it is the probabilistic
propositions themselves that represent the coveted universals?

I have already hinted at the answer I would give, at least until some more
work has been done. I don't believe in 'essentialism' here more than in any
other domain. For me, the whole question of translation universals is not one
of *existence* – 'in the world', so to speak – but one of *explanatory power*. I truly
believe this is one of the most powerful tools we have had so far for going
beyond the individual and the norm-governed, and therefore I will stick to it;
at least for the time being. As a tool, that is, even though not necessarily under
the title of 'universals'.

It so happens that I did use the word 'universals' (well, its Hebrew coun-
terpart) in my 1976 dissertation, but dropped it right away and refrained from
using it ever since, not even when other scholars started using the term (e.g.
Newman 1985–1986). As of the early 1980s, the notion I favored was that of
'laws', and I can see no reason to reverse that decision. In fact, I decided to use
'universals' here mainly because this was the term used by the organizers of the
Workshop (as well as the session in the EST Congress mentioned above) and
most of the participants.

The reason why I prefer 'laws' is not merely because, unlike 'universals', this
notion has the possibility of *exception* built into it (which is important from the
probabilistic point of view because no probability is ever 1), but mainly because
it should always be possible to explain away [seeming] exceptions to a law with
the help of *another* law, operating on *another* level.[11] In brief, I don't believe

the gain in retaining the notion of 'universals' in Translation Studies is worth the price we would be paying for it. But maybe I am wrong. Maybe future work will make me change my mind.

Notes

* A shorter version, focusing on slightly different issues, was presented at the 3rd Congress of EST: "Translation Studies: Claims, Changes and Challenges", Copenhagen, August–September 2001 (forthcoming).

1. I will however highlight possible universals by using SMALL CAPS.

2. Examples of such studies would include Hans Lindquist's account of English adverbs in Swedish translation, on the basis of some 2000 expressions, the first 200 adverbials in each one of ten texts (Lindquist 1984), or Uwe Kjär's doctoral dissertation on the translation into Swedish of German verb metaphors of the type *Der Schrank seufzt*, 1188 in number, occurring in some 4,000 pages of modern German novels (Kjär 1988).

3. Actually, the first links may be missing too: there is no need to assume that words are the lowest-level items that can be observed and submitted to study within Translation Studies, nor that lower-level items are necessarily less interesting objects for study.

4. Incidentally, this account also highlights some of the limitations of corpus studies in their present application to translation: it is [relatively] easy to collect in a fully automatic way immense amounts of material on the lower levels, making the calculation of factual frequencies quite easy and reliable. It becomes more and more complicated, and less and less automatic, the higher one goes up the generality scale, which renders probabilities much more difficult to assess.

5. It is therefore quite surprising that he does not apply a similar approach to translation in the few articles he published about it (e.g. Halliday 1993a).

6. See Note 1.

7. Thus, it is not even agreed whether prototypical translation should be ascribed to 'professionals' (e.g. Halverson 2000) or to 'natural translators' (Harris 1978), nor is there any agreement as to what each one of them means.

8. This argument was developed in Toury 2002.

9. The notion of 'shift' itself will be kept intuitive. The issue of how, and in respect to what, a shift may be discerned and/or measured is controversial and tackling it is bound to take us way off course.

10. This statement itself may well be another candidate for universal-ship.

11. For instance, an expected phonetic change that doesn't occur (which is always a possibility) is often justified as an evidence of having been created at a later period, when the law had stopped being active, or as an evidence of having been imported from without, in a situation of language contact, or as a result of a combination of the two.

References

Halliday, Michael A. K. (1991a). Towards Probabilistic Interpretations. In E. Ventola (Ed.), *Functional and Systemic Linguistics: Approaches and Uses* (pp. 39–61). Berlin/New York: Mouton de Gruyter.

Halliday, Michael A. K. (1991b). Corpus Studies and Probabilistic Grammar. In K. Aijimer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 30–43). Harlow: Longman.

Halliday, Michael A. K. (1993a). Language Theory and Translation Practice. *Rivista internazionale di tecnica della traduzione (Trieste), 0* 15–25.

Halliday, Michael A. K. (1993b). Quantitative Studies and Probabilities in Grammar. In M. Hoey (Ed.), *Data, Description, Discourse: Papers on English Language in Honour of John McH. Sinclaire on his sixtieth birthday* (pp. 1–25). London: Harper Collins.

Halverson, Sandra (2000). Prototype Effects in the 'Translation' Category. In A. Chesterman, N. Gallardo San Salvador, & Y. Gambier (Eds.), *Translation in Context: Selected Contributions from the EST Congress, Granada 1998* (pp. 3–16). Amsterdam/ Philadelphia: John Benjamins.

Harris, Brian (1978). The Difference between Natural and Professional Translation. *Canadian Modern Language Review, 34*, 417–427.

Kjär, Uwe (1988). *Der Schrank seufzt: Metaphern im Bereich des Verbs und ihre Übersetzung.* Göteborg: Acta Universitatis Gothoburgensis.

Lindqvist, Hans (1984). The Use of Corpus-based Studies in the Preparation of Handbooks for Translators. In W. Wilss & G. Thome (Eds.), *Translation Theory and Its Implementation in the Teaching of Translating and Interpreting* (pp. 260–270). Tübingen: Gunter Narr.

Newman, Aryeh (1985–1986). Translation Universals – Perspectives and Explorations. In M. Gaddis Rose (Ed.), *Translation Perspectives* III: *Selected Papers*, *1985–1986* (pp. 69-83). Binghamton: State University of New York at Binghamton.

Reichenbach, Hans (1948). *The Theory of Probability: an Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability.* Berkeley: University of California Press.

Steiner, Erich (2001). *How (Translated and Otherwise Interlingual) Texts Work is Our Way into What, Why and To What Effects. Target, 13*, 343–348.

Toury, Gideon (1977). *Translational Norms and Literary Translation into Hebrew, 1930–1945.* Tel Aviv: The Porter Institute for Poetics and Semiotics, Tel Aviv University. [A slightly revised version of a 1976 PhD Dissertation; in Hebrew.]

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

Toury, Gideon (1987). Integrating the Cultural Dimension into Translation Studies: An Introduction. In G. Toury (Ed.), *Translation across Cultures* (pp. 1–8). New Delhi: Bahri.

Toury, Gideon (1993). Probabilistic Explanations in Translation Studies. Plenary Lecture in the Postgraduate Seminar Translation Theory and Research, Savonlinna School of Translation Studies, University of Joensuu.

Toury, Gideon (1995). *Descriptive Translation Studies and beyond.* Amsterdam: John Benjamins.

Toury, Gideon (2002). What Is the Problem with 'Translation Problem'? In B. Lewandowska-Tomaszczyk & M. Thelen (Eds.), *Translation and Meaning*, Part 6 (pp. 57–71). Maastricht: School of Translation and Interpreting.

Toury, Gideon (Forthcoming). Probabilistic Explanations in Translation Studies: Universals – Or a Challenge to the Very Concept? In G. Hansen, K. Makmkjær, & D. Gile (Eds.), *Translation Studies: Claims, Changes and Challenges: Proceedings of the 3rd Congress of EST*. Amsterdam/Philadelphia: John Benjamins.

Beyond the particular*

Andrew Chesterman
University of Helsinki

Translation scholars have proposed and sought generalizations about translation from various perspectives. This paper discusses three main ways of getting "beyond the particular": traditional prescriptive statements, traditional critical statements, and the contemporary search for universals in corpus studies. There are a number of problems with each of the approaches.

1. Introduction

Any science seeks generalities. The aim is to transcend knowledge of particular cases by discovering general regularities or laws, or by proposing general descriptive hypotheses that cover more than a single case. Only by looking for similarities between single cases, and then generalizing from these, can a science progress to the ability to make predictions concerning future or unstudied cases. Only in this way can any discipline progress towards an understanding of the general explanatory laws that are relevant in its field. And only in this way can a discipline create links with neighbouring disciplines. An interdiscipline like Translation Studies will be doomed to stagnation if this striving towards the general is neglected.

Seeking generalities means looking for similarities, regularities, patterns, that are shared between particular cases or groups of cases. Such a search does not deny the existence or importance of that which is unique in each particular case; nor does it deny the existence or importance of differences between cases. At its best, such research allows us to see both similarities and differences in a perspective that increases our understanding of the whole picture, and also of how this picture relates to other pictures.

Translation Studies has sought to escape the bounds of the particular in three ways. All three routes have meant looking at (and for) linguistic features which relate translations to (1) the source text and (2) the target language. I

will refer to these routes as (a) the prescriptive route, (b) the pejorative route, and (c) the descriptive route. Along the prescriptive route we find statements about various features which all translations, or all translations of a given sort, should or should not manifest, ideally. Along the pejorative route we find statements about undesirable features which all, or most, or some type of, translations are thought to manifest, in reality. Along the descriptive route we find statements about possible universal features of translations or subsets of translations, without overt value judgements.

Each route has its problems, and each has made contributions.

2.   The prescriptive route

The oldest, traditional route away from the particular has been the stating of prescriptive generalities that purport to hold for all translations. These statements typically have the form: "All translations should have feature X / should not have feature Y", and thus reflect some kind of translation ideal, universally valid. Examples abound in the early literature: Dolet's and Tytler's translation principles, for instance. The culmination of this route is perhaps reached in Savory's famously paradoxical list of mutually contradictory principles.

Dolet (La manière de bien traduire d'une langue en aultre, 1540; three of his five general principles)

Translations should not be word-for-word renderings of the original.
Translations should avoid unusual words and expressions.
Translations should be elegant, not clumsy.

Tytler (Essay on the principles of translation, 1797)

Translations should give a complete transcript of the ideas of the original.
Translations should be in the same style as their source texts.
Translations should be as natural as original texts.

Savory (1968: 54)

1. A translation must give the words of the original.
2. A translation must give the ideas of the original.
3. A translation should read like an original work.
4. A translation should read like a translation.
5. A translation should reflect the style of the original.

6. A translation should possess the style of the translation.
7. A translation should read as a contemporary of the original.
8. A translation should read as a contemporary of the translation.
9. A translation may add to or omit from the original.
10. A translation may never add to or omit from the original.
11. A translation of verse should be in prose.
12. A translation of verse should be in verse.

Problem: overgeneralization (neglect of differences). The weakness of this route is of course that no account is here taken of the fact that translations are not all of a kind: some prescriptive principles may be valid for some types of translation (or types of text) and other principles for other types. As soon as this is realized, the need arises for a translation typology.

Perhaps the first attempt to make such a typology was that of Jerome, who claimed as follows:

> Jerome (De optimo genere interpretandi, 395)
>
> Translations of sacred texts must be literal, word-for-word (because even the word order of the original is a holy mystery and the translator cannot risk heresy).
> Translations of other kinds of texts should be done sense-for-sense, more freely (because a literal translation would often sound absurd).

Problem: fallacy of converse accident. This is the fallacy of generalizing from a non-typical particular. Here again, differences are neglected. What we find is that statements based on translating a particular kind of text, such as a literary text or the Bible, are assumed to hold good for all kinds of texts – and indeed all kinds of translations. Traces of this fallacy are to be found in quite recent publications on translation theory. A well-known anthology of essays that came out in 1992 was entitled "Theories of Translation" (edited by Schulte and Biguenet). Most of the essays are indeed classics. But all except two deal exclusively with literary translation. The impression is given that translation theory can be more or less equated with literary translation theory – as if literary translation was typical of all translation. A similar impression is given by Venuti's recent collection of readings (2000), the great majority of which concern literary translation.

Problem: idealization. By this I mean the evident underlying belief in perfection, in a perfect translation that would be absolutely equivalent and also ab-

solutely natural. The influence of theological myths is strong here, such as that of the 72 translators of the Septuagint who all arrived miraculously at the same solutions...

Contribution: first attempts to generalize.   These early prescriptive statements were at least a first attempt to get beyond the particular, to establish more general principles and parameters. The statements were based on implicit predictive hypotheses based on the following argument:

–   A given translation X is a good translation (i.e. this is someone's reaction to it, its effect on their judgement).
–   This quality judgement is based on the presence of features ABC in the translation X.
–   Therefore, all translations with features ABC will be good, people will react to them in this way.

The argument only works, of course, if we accept three assumptions: that the quality assessment of translation X really is caused by the presence of features ABC and not something else; that all translations are of the same type as X; and that features ABC are universal indicators of high quality.

Contribution: subsequent attempts at typologies.   Since Jerome, there have been many attempts to set up typologies of translation (see e.g. Chesterman 1999 for a brief survey). None have yet become generally accepted.

Contribution: concern with translation quality.   Quality is a central concern of all those who are involved in the practical work of translation. The descriptivists have perhaps over-reacted against traditional prescriptivism in their desire to place Translation Studies on a more scientific basis. However, if quality assessments are seen as part of the effects that a translation has, they need not be excluded from empirical analysis. Defining quality, and devising reliable measures of it, are genuine research problems that should form part of research into translation effects.

3.   The pejorative route

The second route away from the particular is related to the first, but takes a different direction. Here, all translations (or: all translations of a certain kind) are regarded as being deficient in some way. That is, an attempt is made to

characterize a set of translations in terms of certain negative features. Along this route we find the traditional tropes of loss and betrayal, the view of translations as merely secondary texts, as necessarily either not faithful or not beautiful.

They are not faithful because they are too free, too fluent, too naturalized, too domesticated: these deficiencies are often noted by literary critics. Translations are not beautiful if they contain unnatural target language, such as that frequently noticed in tourist brochures, menus, etc. (For dozens of examples, surf the web for Tourist English.)

Along this pejorative path, we find hundreds of statements to the effect that translators are doomed to eternal failure, they are objects of scorn or laughter. The literature abounds in critics' lists of typical translation weaknesses. One of the most recent examples is represented by Antoine Berman, with respect to literary translation: in brief, he claims that these are typically too free. Here is his list of the "deforming tendencies" of literary translation (Berman 1985; see also Munday 2001: 149–151).

– Rationalization (making more coherent)
– Clarification (explicitation)
– Expansion
– Ennoblement (more elegant style)
– Qualitative impoverishment (flatter style)
– Quantitative impoverishment (loss of lexical variation)
– Destruction of rhythms
– Destruction of underlying networks of signification
– Destruction of linguistic patternings (more homogeneous)
– Destruction of vernacular networks or their exoticization (dialect loss or highlighting)
– Destruction of expressions and idioms (should not be replaced by TL equivalent idioms)
– Effacement of the superimposition of languages (multilingual source texts)

A similar line of argument is to be found in Kundera's ideas about translation, particularly the translations of his own works (Kundera 1993: 123f.). He complains about the way translators violate metaphors, seek to enrich simple vocabulary, reduce repetition, spoil sentence rhythms by altering punctuation, even change the typography.

Some of these putative deficiencies reoccur in the descriptive work we shall come to below.

Problem: assumptions about quality – overgeneralization again. The weakness of this kind of approach is not so much a failure to develop a translation typology; rather, it is a very restricted *a priori* view of what constitutes an acceptable translation in the first place. This view is so narrow that a great many translations are automatically criticized, although they might be perfectly acceptable according to other criteria than those selected by the critic in question, e.g. relating to strict formal equivalence or flawless target language. After all, not all translations need to be perfectly natural TL. (By "natural" here I mean 'unmarked', in the sense that readers typically do not react *de dicto*, to the linguistic form itself.) We usually understand the funny menus and notices – they are often part of the amusement of a holiday, we may even expect them. And unnatural (marked) language will be less noticed by non-natives anyway. With respect to the alleged weaknesses of much literary translation, one can point out that most readers of literary translations may well prefer a freer, more natural version. The criticism may boil down to no more than personal preference.

Problem: assumption of the universality of formal stylistic universals. This is a different kind of problem. The literary critics I referred to above seem to overlook the fact that a given formal feature (repetition, say) may have quite different effects on readers in different cultures, where there may be quite different rhetorical and stylistic norms. These critics thus neglect the possibility of cultural relativity, in favour of a belief in form for form's sake, a belief in the existence, distribution and frequency of formal stylistic universals that have yet to be demonstrated. Formal equivalence is valued, dynamic equivalence is not.

Problem: socio-cultural effect on translator status. One highly undesirable effect of these pejorative generalizations is of course the depressing impact it has on the public perception of the translator's role, and indeed on translators' own perception of themselves, as poor creatures doomed to sin.

Contribution: concern with quality. These pejorative views do nevertheless reveal a concern with translation quality, albeit narrowly understood. From this route away from the particular we learn the need to develop more sophisticated and varied criteria for assessing translation quality. (For a recent selection of views on quality assessment, see Schäffner 1998 and the special issue of *The Translator* 6 (2), 2000.)

Contribution: awareness of ethical issues. Another contribution worth mentioning is the way in which critics such as Berman foreground issues concerning ethnocentrism and more generally the representation of the Other. This helps us to see the wider philosophical context in which translation takes place, and has fuelled quite a bit of later research on translation ethics (see the special issue of *The Translator* 7 (1), 2001).

4. The descriptive route

The third route away from the particular is represented by recent corpus-based work on what some call translation universals. One of the origins of such work has been Frawley's notion (1984) of translations as constituting a third code, distinct from the source-language and target-language codes. Another origin has been hypotheses like that of Blum-Kulka (1986) on explicitation, and yet another has been Toury's (1995) proposals about translation laws. We should also mention the background of work in linguistics on language universals, and in sociolinguistics on language variation.

Progress along this descriptive route seems to be moving along two roads simultaneously: the high road and the low road. On the high road, we find claims that indeed purport to cover all translations, and so they can fairly be said to be claims about universal features. These claims fall into two classes, corresponding to the two contrastive textlinguistic relations that form the core of linguistic research on translation: the equivalence relation with the source text, and the relation of textual fit with comparable non-translated texts in the target language. In other words, use is made of two different reference corpora. Some hypotheses claim to capture universal differences between translations and their source texts, i.e. characteristics of the way in which translators process the source text; I call these *S-universals* (S for source). Others make claims about universal differences between translations and comparable non-translated texts, i.e. characteristics of the way translators use the target language; I call these *T-universals* (T for target). T-universals are the descriptive equivalent to the criticisms of unnaturalness, of translationese, made in the pejorative approach.

Below are some examples of both types of proposed universals. Note that these claims are hypotheses only; some have been corroborated more than others, and some tests have produced contrary evidence, so in most cases the jury is still out.

Potential S-universals

– Lengthening: translations tend to be longer than their source texts
   (cf. Berman's expansion; also Vinay & Darbelnet 1958: 185; et al.)
– The law of interference (Toury 1995)
– The law of standardization (Toury 1995)
– Dialect normalization (Englund Dimitrova 1997)
– Reduction of complex narrative voices (Taivalkoski 2002)
– The explicitation hypothesis (Blum-Kulka 1986, Klaudy
   1996, Øverås 1998) (e.g. there is more cohesion in translations)
– Sanitization (Kenny 1998) (more conventional collocations)
– The retranslation hypothesis (later translations tend to be
   closer to the source text; see *Palimpsestes* 4, 1990)
– Reduction of repetition (Baker 1993)

Potential T-universals

– Simplification (Laviosa-Braithwaite 1996)

      Less lexical variety
      Lower lexical density
      More use of high-frequency items

– Conventionalization (Baker 1993)
– Untypical lexical patterning (and less stable) (Mauranen 2000)
– Under-representation of TL-specific items (Tirkkonen-Condit 2000)

Research then proceeds by operationalizing these general claims, i.e. interpreting them in concrete terms, and then testing them on various kinds of data in order to see how universal they actually are. Do they, for instance, apply to some subset of all translations rather than the total set? This leads us to consider the second direction pursued by modern descriptive research, the low road.

Here, research moves in more modest steps, generalizing more gradually away from particular cases towards claims applying to a group of cases, then perhaps to a wider group, and so on. The movement is bottom-up (starting with the particular) rather than top-down (starting with the general). True, a universal hypothesis might also be tentatively proposed on the basis of empirical results pertaining only to a subset. Subset generalizations fall into the same two classes as the universal claims mentioned above: claims about the source/target relation, and claims about the translated/non-translated relation.

A crucial point in this bottom-up approach is the criteria on which the subset is defined. These criteria in effect define the conditions that determine and limit the scope of the claim. Several have been used, either separately or in combination, such as the following:

– *Language-bound criteria*: claims pertain to translations between a given language pair and a given translation direction. See e.g. classics like Vinay and Darbelnet (1958) on French and English, Malblanc (1963) on French and German. The results of traditional contrastive analysis and contrastive rhetoric come in useful here, at the explanatory level, when we look for the language-bound causes of translation features (e.g. Doherty 1996). Maia (1998), for instance, considers features of English word order that appear to affect Portuguese word order in translations from English: these translations show a different distribution of word order variants from that found in non-translated Portuguese texts.
– *Time-bound and place-bound criteria*: claims pertain to a particular period, in a particular culture. See e.g. Toury (1995:113f.) on early 20th-century Hebrew norms for poetry translation.
– *Type-bound criteria*: claims pertain to a particular type of translation (characterized e.g. by a given text-type or skopos-type). Many examples: Bible translation, subtitling, technical, poetry, comic strips, gist translation... E.g. Mauranen (2000) found that translations of popular non-fiction deviated more from lexical patterning norms than did translations of academic texts.
– *Translator-bound criteria*: claims pertain to translations done by a particular translator. See e.g. Baker (2000) on translators' individual style). Or they pertain to translators of a particular kind (trainees; men/women; to L1 or L2; ...).
– *Situation-bound criteria*: claims pertain to particular conditions of the publishing or editorial process, in-house stylistic conventions and the like. E.g. Milton (2001).

In this kind of research, we might find that given features are typical (or not typical) of some subset of translations; or that given features seem to be typical (or not typical) of more than one subset.

 This third, descriptive, route away from the particular is not without its problems, either. Indeed, some scholars have preferred to reject this route altogether and restrict their attention to what makes any given translation unique, rather than focus on its similarities with other translations.

Problem: testing. Tests of these claims sometimes produce confirmatory evidence, sometimes not. But how rigorous are the tests? If you are investigating, say, explicitation or standardization, you can usually find *some* evidence of it in any translation; but how meaningful is such a finding? It would be more challenging to propose and test generalizations about *what* is explicitated or standardized, under what circumstances, and test those. To find *no* evidence of explicitation or standardization would be a surprising and therefore strong result. Stronger still would be confirmation in a predictive classification test, as follows (based on a suggestion by Emma Wagner, personal communication, 2001). If these universals are supposed to be distinctive features of translations, they can presumably be used to identify translations. So you could take pairs of source and target texts, and see whether an analysis of some S-universal features allows you to predict which text in each pair is the source and which the target text. For each pair you would have to do the analysis in two directions, assuming that each text in turn is source and target, to see which direction supports a given universal tendency best. Or you could take a mixed set of texts consisting of translations and non-translations and analyse them for a given T-universal feature, and use the results to predict the category assignment of each text (= translation or not). Some universals might turn out to be much more accurate predictors than others.

Problem: representativeness. Since we can never study all translations, nor even all translations of a certain type, we must take a sample. The more representative the sample, the more confidence we can have that our results and claims are valid more generally. Measuring representativeness is easier if we have access to large machine-readable corpora, but there always remains a degree of doubt. Our data may still be biased in some way that we have not realized. This is often the case with non-translated texts that are selected as a reference corpus. Representativeness is an even more fundamental problem with respect to the translation part of a comparable corpus. It is not *a priori* obvious what we should count as corpus-valid translations in the first place: there is not only the tricky borderline with adaptations etc., but also the issue of including or excluding non-professional translations or non-native translations, and even defining what a professional translation is (see Halverson 1998). Should we even include "bad" translations? They too are translations, of a kind.

Problem: universality. Claims may be made that a given feature is universal, but sometimes the data may only warrant a subset claim, if the data are not

representative of all translations. Many "universal" claims have been made that actually seem to pertain only to literary or to Bible translation. More fundamentally, though: since we can ever only study a subset of all translations past and present, there is always the risk that our results will be culture-bound rather than truly universal (Tymoczko 1998). Concepts of translation itself are culture-bound, for a start; even prototype concepts may be, too. We can perhaps never totally escape the limits of our own culture-boundness, even if this might be extended e.g. to a general "Western culture". This means that claims of universality can perhaps never be truly universal.

In the light of these problems and reservations, it is obvious that any claim about a translation universal can really only be an approximation. But this does not matter, as long as scholars are aware of what they are claiming. After all, what these corpus scholars are basically doing is seeking generalizations. We seek generalizations that are as extensive as possible. Less-than-universal claims can still be interesting and valuable. Any level of generalization can increase understanding.

Problem: conceptualization and terminology. Here there is still a great deal to be clarified. I made one proposal above, about distinguishing between S- and T-universals. Baker's original use (1993) of the term "universal" seems to have to refer to T-universals, since her point of comparison is non-translated, original texts; however, several of the examples of previous research that she mentions are based on evidence from a comparison with source texts, and hence concern S-universals (such as the reduction of repetition). If your corpus does not actually contain source texts, you surely cannot study S-universals. Other scholars have, however, used the term to apply either to S-universals alone, or more generally to both S and T types. I think that the use of the term "universal" itself is valid and useful, provided it is kept for claims that are actually hypothesized to be universal, not specific to some subset of translations.

Some scholars prefer to refer to these claims as hypotheses, such as the explicitation hypothesis (Blum-Kulka and others) or the simplification hypothesis (Laviosa-Braithwaite 1996), or the retranslation hypothesis. Others speak of laws: cf. Toury's proposed laws of interference and standardization. Chevalier (1995) writes about "figures of translation", comparable to rhetorical figures; the occurrence of these figures is contrasted with translation alternatives that are more neutral or natural or "orthonymic", in the same way that in rhetorical analysis one can distinguish between utterances with or without rhetorical

embellishment. Still other scholars prefer to look for core patterns, or simply widespread regularities.

Claims about universals are in fact examples of descriptive hypotheses – unrestricted descriptive hypotheses, with no scope conditions. As soon as we limit the scope of the claims to some subset of translations, we are proposing restricted descriptive hypotheses.

When it comes to the hypotheses themselves we find a plethora of terms that appear at first sight to mean more or less the same thing (e.g. standardization, simplification, levelling, normalization, conventionalization). Sometimes these are used to refer to a feature of difference between translations and their source texts, and sometimes to a feature of difference between translations and non-translated texts. These latter are called 'parallel' texts by some scholars, 'comparable' texts by others, and 'original' texts by still others. I now use 'non-translated' to avoid confusion: this also gives the convenient abbreviation NT, to go with ST and TT.

And further: some of the terms appear to be ambiguous between a process reading (from source text to translation) and a product reading: e.g. those ending in *-tion* in English. We do need to standardize our terminology here.

Problem: operationalization.   Different scholars often operationalize these abstract notions in different ways – which again makes it difficult to compare research results. We need more replication, and this means explicit descriptions of methodology.

Problem: causality.   A final major problem has to do with causality and how to study it. To claim that a given linguistic feature is universal is one thing. But we would also like to know its cause or causes. Here, we can currently do little more than speculate as rationally as possible. The immediate causes of whatever universals there may be must be sought in human cognition – to be precise, in the kind of cognitive processing that produces translations. Translations arise, after all, in the minds of translators, under certain causal constraints. One source of these constraints is the source text, or rather its meaning or intended message. The translator is constrained by "what was said" in the earlier text. More precisely, translators are constrained by what they *understand* was said in the source text. This inevitable interpretation process acts as a filter; and it is this filtering that seems to offer a site for the explanation for some of the S-universals that have been claimed, such as those concerning standardization and explicitation. Filtering involves reducing the irrelevant

or unclear, purifying, selecting the essence. How it works in detail remains to be seen.

Constraints on cognitive processing in translation may also be present in other kinds of constrained communication, such as communicating in a non-native language or under special channel restrictions, or any form of communication that involves relaying messages, such as reporting discourse, even journalism. It may be problematic, eventually, to differentiate factors that are pertinent to translation in particular from those that are pertinent to constrained communication in general.

Other kinds of explanations may be sought in the nature of translation as a communicative act, and in translators' awareness of their socio-cultural role as mediators of messages for new readers (see e.g. Klaudy 1996). Translators tend to want to reduce entropy, to increase orderliness. They tend to want to write clearly, insofar as the skopos allows, because they can easily see their role metaphorically as shedding light on an original text that is obscure – usually unreadable in fact – to their target readers: hence the need for a translation. Their conception of their role may give a prominent position to the future readers of their texts; this may have been emphasized in their training, for instance. It is this conception of their mediating role that may offer some explanation for the tendency towards explicitation, towards simplification, and towards reducing what is thought to be unnecessary repetition – to save the readers' processing effort. In terms of relevance theory (which defines relevance as the optimum cost-benefit ratio between processing effort needed and cognitive effect produced – see Gutt 2000), translators as a profession are perhaps more aware than other writers of the cost side of the relevance equation. It may be that translators see explicitation in some sense as a norm; perhaps it was even presented as such in their training.

This raises the interesting question of whether there might exist universal norms of communication which could provide explanatory principles for possible translation universals, perhaps along the lines of Grice's maxims (Grice 1975) or notions of politeness. However, these will have to be modified somewhat if they are to be made appropriate to non-Anglo-Saxon cultures.

Research into the *effects* caused by potential universals is still in its infancy. Effects on readers, on translator trainers, and on translators themselves would all be worth studying. It may be that the more we know about T-universals, for instance, the more scholars or trainers will be tempted to see them as undesirable features that should be avoided – at least in translations whose skopos includes optimum naturalness. On the other hand, as the sheer quantity of translations grows and target-language norms become blurred, it may be

that readers will become more tolerant of apparent non-nativeness; different cultures might differ considerably in this respect. One long-term effect of knowledge about S-universals on source-text writers might even be a greater concern for the clarity of the source text, in order to facilitate the translator's task and lessen the need for explicitation. This in turn could lead to greater fidelity to the original.

Contribution: methodological. The prime benefit so far of this kind of descriptive research has, I think, been methodological. Corpus-based research into translation universals has been one of the most important methodological advances in Translation Studies during the past decade or so, in that it has encouraged researchers to adopt standard scientific methods of hypothesis generation and testing. This kind of research also makes it obvious that we need to compare research results across studies and take more account of what others have done. The application of methods from corpus linguistics has encouraged more use of quantitative research. Research on descriptive hypotheses has also brought new knowledge about translation, and a host of new hypotheses to be tested. It has thus helped to push Translation Studies in a more empirical direction.

Contribution: interdisciplinarity. Another benefit has been the highlighting of interdisciplinarity. Descriptive research on universals shows how Translation Studies must be linked to other fields, not only within linguistics but within the human sciences more generally (cognitive science, for example, and cultural anthropology).

Contribution: concern with translation quality. Perhaps paradoxically, this descriptive approach has also drawn our attention to subtle aspects of text and translation quality. There are many potential applications here: translators who are aware of these general tendencies (even if they may not be universal ones) can choose to resist them. Non-native translators can make good use of quantitative information, banks of comparable non-translated texts, to make their own use of the target language more natural, and they can run tests to check the naturalness of aspects of their translations. This facility may lead to the gradual blurring of the distinction between native and non-native translators at the professional level, which in turn should have an influence on assumptions held by many translation theorists about the exclusive status of translation into the native language. (This issue is discussed e.g. in Campbell 1998 and Pokorn 2000.)

What we need, therefore is much more replicated work on testing different restricted and unrestricted hypotheses on different corpora. We need to standardize our main concepts and ways of operationalizing, for greater research cooperation. We need to relate descriptive hypotheses to each other, at still more abstract levels. We need to develop electronic corpus tools. We need to generate new descriptive hypotheses. And we need to work on testable explanatory hypotheses in order to account for the evidence we find.

Note

* This article is based on three conference presentations, during which my ideas on the topic have developed. One paper was read at the Third EST Congress in Copenhagen in August-September 2001, as part of the session on universals; another was read at the Symposium on Contrastive Analysis and Linguistic Theory at Ghent in September 2001; and a third at the Conference on Universals at Savonlinna in October 2001. There is some overlap between the published versions of all three presentations. I am grateful for all the critical comments and feedback I have had at these meetings.

References

Baker, Mona (1993). Corpus linguistics and Translation Studies: Implications and applications. In M. Baker et al. (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). Amsterdam: John Benjamins.

Baker, Mona (2000). Towards a methodology for investigating the style of a literary translator. *Target, 12*(2), 241–266.

Berman, Antoine (1985). *Traduction et la lettre ou l'auberge du lointain.* Paris: Seuil.

Blum-Kulka, Shoshana (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17–35). Tübingen: Narr.

Campbell, Stuart (1998). *Translation into the Second Language.* London: Longman.

Chesterman, Andrew (1999). Translation typology. In A. Veisbergs & I. Zauberga (Eds.), *The Second Riga Symposium on Pragmatic Aspects of Translation* (pp. 49–62). Riga: University of Latvia.

Chevalier, Jean-Claude (1995). D'une figure de traduction : le changement de 'sujet'. In J-C. Chevalier & M-F. Delport, *L'Horlogerie de Saint Jérôme* (pp. 27–44). Paris: L'Harmattan.

Doherty, Monika (1996). *Information Structure: a key concept for translation theory. Linguistics, 34*(3). (Special issue.)

Dolet, Estienne (1540). *La Manière de Bien Traduire d'une Langue en Aultre.* Paris: Marnef.

Englund Dimitrova, Birgitta (1997). Translation of dialect in fictional prose – Vilhelm Moberg in Russian and English as a case in point. In J. Falk, G. Magnusson, G. Melchers, & B. Nilsson (Eds.), *Norm, Variation and Change in Language. Proceedings of the centenary meeting of the Nyfilologiska sällskapet, Nedre Manilla 22–23 March, 1996* (pp. 49–65). Stockholm: Almqvist and Wiksell.

Frawley, William (1984). Prolegomenon to a theory of translation. In W. Frawley (Ed.), *Translation. Literary, Linguistic and Philosophical Perspectives* (pp. 159–175). Newark: University of Delaware Press.

Grice, Paul (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41–58). New York: Academic Press.

Gutt, Ernst-August (2000). *Translation and relevance. Cognition and context*. Revised edition. Manchester: St. Jerome Publishing.

Halverson, Sandra (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.

Jerome, Eusebius Hieronymus [395] (1997). De optime genere interpretandi. Translated by P. Carroll as On the best kind of translator. In D. Robinson (Ed.), *Western Translation Theory from Herodotus to Nietzsche* (pp. 22–30). Manchester: St. Jerome Publishing.

Kenny, Dorothy (1998). Creatures of habit? What translators usually do with words. *Meta, 43*(4), 515–523.

Klaudy, Kinga (1996). Back-translation as a tool for detecting explicitation strategies in translation. In K. Klaudy et al. (Eds.), *Translation Studies in Hungary* (pp. 99–114). Budapest: Scholastica.

Kundera, Milan (1993). *Les Testaments Trahis*. Paris: Gallimard.

Laviosa-Braithwaite, Sara (1996). *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*. Unpublished PhD thesis, UMIST, Manchester.

Maia, Belinda (1998). Word order and the first person singular in Portuguese and English. *Meta, 43*(4), 589–601.

Malblanc, Alfred (1963). *Stylistique Comparée du Français et de l'Allemand*. 2nd edition. Paris: Didier.

Mauranen, Anna (2000). Strange strings in translated language. A study on corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects* (pp. 119–141). Manchester: St. Jerome Publishing.

Milton, John (2001). The figure of the factory translator. Paper presented at the Third EST Congress, August 30 – September 1, 2001, Copenhagen.

Munday, Jeremy (2001). *Introducing Translation Studies. Theories and Applications*. London: Routledge.

Øverås, Linn (1998). In search of the third code: an investigation of norms in literary translation. *Meta, 43*(4), 571–588.

Palimpsestes 4, 1990.

Pokorn, Nike Kocijancic (2000). Translation into a non-mother tongue in translation theory: deconstruction of the traditional. In A. Chesterman et al. (Eds.), *Translation in Context* (pp. 61–72). Amsterdam: John Benjamins.

Savory, Theodore (1957/1968). *The Art of Translation*. London: Cape.

Schulte, Rainer, & Biguenet, John (1992). *Theories of Translation*. Chicago: University of Chicago Press.

Schäffner, Christina (1998). *Translation and Quality*. Clevedon: Multilingual Matters.

Taivalkoski, Kristiina (2002). Traduire la mixité formelle: l'exemple des premières (re)traductions de Fielding en France. *Faits de Langue, 19*, 85–97.

Tirkkonen-Condit, Sonja (2000). In search of translation universals: non-equivalence or 'unique' items in a corpus test. Paper presented at the UMIST/UCL Research Models in Translation Studies Conference, 28–30 April 2000, Manchester.

Toury, Gideon (1995). *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.

The Translator, 6(2), 2000.

The Translator, 7(1), 2001.

Tymoczko, Maria (1998). Computerized corpora and the future of Translation Studies. *Meta, 43*(4), 653–659.

Tytler, Alexander Fraser [1797] (1978 [reprint]). *Essay on the Principles of Translation*. Amsterdam: Benjamins.

Venuti, Lawrence (2000). *The Translation Studies Reader*. London: Routledge.

Vinay, Jean-Paul and Darbelnet, Jean (1958). *Stylistique Comparée du Français et de l'Anglais*. Paris: Didier.

When is a universal not a universal?

Some limits of current corpus-based methodologies for the investigation of translation universals

Silvia Bernardini and Federico Zanettin
University of Bologna / University for Foreigners of Perugia

This paper raises a number of concerns relating to the notion of universality in translation and to the methodology adopted in the search for translation "universals". The term itself, it is suggested, may be misleading if applied to corpus-based research, where the emphasis should be *first* on the relations between translated texts and the socio-cultural constraints under which they were produced and *then* on the cognitive processes underlying translation activities. Taking examples from the CEXI corpus, a parallel bi-directional corpus of English and Italian currently under construction, the paper illustrates the working of some such constraints on corpus design. With reference to non-fiction texts, it shows how two different cultures (Italy vs. the U.S.) reciprocally select for translation texts belonging to different textual typologies, resulting in the possibility of skewed distributions within comparable corpora. Similarly, with reference to fiction texts, it shows how Italian texts translated into English tend to be canonical high-brow ones, whilst this is not the case with English texts translated into Italian. We suggest that the effect of such contextual variables over translation strategies and norms should not be neglected in translation research. One suggestion in this direction is to set up corpus resources so as to allow multiple comparisons across subcorpora, such that each component can be used as a control for the mirror one.

1. Introduction: universals and DTS

The current fascination with universals in (corpus-based) descriptive translation studies (DTS) may appear somewhat surprising. A research methodology with a double lineage, corpus-based DTS has inherited the Firthian/Hymesian

views of linguistics as the study of language in use underlying corpus linguistics, as well as DTS views of translation research as the target-oriented, situationally-constrained study of translation practices:

> [...] a normal child acquires knowledge of sentences, not only as grammatical, but also as appropriate. He or she acquires competence as to when to speak, when not, and as to what to talk about with whom, when, where, in what manner.   (Hymes 1972: 277)

> Throughout the period of growth we are progressively incorporated into our social organization, and the chief condition and means of that incorporation is learning to say what the other fellow expects us to say under given circumstances. [...] Most of the give-and-take of conversation in our everyday life is stereotyped and very narrowly conditioned by our particular type of culture.
> (Firth 1935: 67, 69)

> 'translatorship' amounts first and foremost to being able to *play a social role*, i.e. to fulfil a function allotted by a community [...] in a way which is deemed appropriate in its own terms of reference.   (Toury 1995: 53)

Quests for "universals" – cognitively-basic, situationally-unconstrained theoretical constructs which lie at the basis of generative/typological approaches to linguistics – would appear to be at odds with the very premises of this approach. Although use of the term *universal* in DTS is generally qualified and accurately glossed to point out that in fact it is a general tendency, or widespread norm, that is postulated, rather than an absolute truth, it is nevertheless true that at least half a century of linguistic research and theorization is attached to the term "universal", and this can hardly be swept under the carpet.

Accordingly, in this paper we shall attempt to steer clear of controversial notions of universality, and aim, in more down-to-earth manner, to shed light on (some) interrelations between parameters of situational and cultural variation and patterns of linguistic usage as observable "in an adequate *corpus inscriptionum*" (Firth 1956: 106). Our major concern here is that of evaluating the adequacy of a corpus in the quest for norms and laws of translational behaviour (Toury 1995: 259–279), as a first, largely methodological step in preparation for more ambitious quests.

We prefer the notion of *law* (as formulated by Toury: *if X, then the greater/the lesser the likelihood that Y* [1995: 256]) to that of *universal*, insofar as laws may be proposed that describe widely – and even universally – followed norms. Unlike universals, however, laws in social science are subject to conditioning factors of various kinds, and as such would appear to be much more

in tune with neo-Firthian linguistics and much more amenable to discovery by means of corpus analyses.

2. Methodological issues in corpus-based DTS

Corpus-based studies of potential universals of translation behaviour have tended to focus on the idea that translated texts as a whole are "simpler" and more "conventional" than both their source texts and "comparable" texts originally produced in the target language. A number of descriptive labels have been proposed in order to account for such phenomena, such as simplification, explicitation, normalization, repetition avoidance, levelling out, disambiguation and standardization (e.g. Baker 1995, 1996; Schmied & Schäffler 1996; Laviosa 1998a, 1998c; Olohan & Baker 2000; Olohan 2001).[1] Electronic corpora make quantitative analyses of these features possible and in some cases relatively straightforward. These data may shed light on choices made unconsciously by translators, providing the researcher with more "objective" data than can be obtained through manual comparisons of single source and target texts.

Within corpus-based DTS, attention has focused mainly on the comparison of translations and original texts in the same language, or "monolingual comparable corpora". The principle behind this approach is that comparison of a corpus of translations with one of non-translations will highlight features of the former, which can be explained in terms of the value added to the text by the translation process. Investigations have used global frequency measures such as type/token ratio and lexical density (defined as the percentage of grammatical to lexical words, Laviosa 1998b: 566), as well as measures relating to particular lexical features and syntactic structures (e.g. Olohan 2001; Olohan & Baker 2000).

On the methodological side, attention has been devoted to the design of monolingual comparable corpora for translation research (see e.g. Laviosa 1997). It has been suggested that in order to eliminate possible source language bias, corpora should include translations from different languages, and that the corpora compared should cover "a similar domain, variety of language and time span, and be of comparable length" (Baker 1995: 234).

The question of "how comparable can comparable corpora be" has long worried researchers, and rightly so, being key to the evaluation of validity and to the replicability of results. One aspect that appears to have been underestimated is the potential bias deriving from the operation of what Toury (1995) calls "preliminary norms" – translation policies affecting, among other

things, the choice of texts to be translated (Even-Zohar 2000 [1978/1990]; Vanderauwera 1985):

> Dutch fiction is chosen for translation either in the function of assumed target taste or in that of the status the work has acquired at the source pole, often as a combination of the two.   (Vanderauwera 1985:132)

In our experience of corpus construction, this issue has been of central importance in assessing the comparability of different corpora.

3.   Issues in translation corpus design and construction: the CEXI example

CEXI is an English-Italian Translational Corpus being developed at the School for Interpreters and Translators, University of Bologna at Forlì. The aim of the project is to arrive at a bi-directional and parallel corpus of approximately four million words of contemporary texts in two languages, XML-tagged following the TEI guidelines (Sperberg-McQueen & Burnard 2001), aligned and accessible online (see Zanettin 2000, 2002). Following projects like the *ENPC* for Norwegian, *Compara* for Portuguese, and the *Chemnitz* corpus for German, CEXI is restricted to what is probably the most prototypical (and most easily samplable) form of translation, i.e. printed books. Altogether 624 titles (half translated from English into Italian, the other half translated from Italian into English) were randomly selected from the *Unesco Index Translationum* database (1998), half of them fiction (roughly corresponding to the Universal Decimal Classification category of Literature/Children's Literature as assigned within the Index Translationum) and half non-fiction (divided into nine subcategories following the Index/UDC criteria). After removing titles which were repeated, outside our time frame, impossible to locate etc., requests for permission were made to the copyright owners.

3.1   Preliminary norms 1: non-fiction

While trying to set up a sampling frame for the non-fiction component of the corpus, it became clear that, in the *Index Translationum* at least, a different "weight" is associated with the different genres in each direction (on this point see also Mauranen 2001). The numbers of texts translated from Italian into English and from English into Italian in each of the UDC subdomains are not comparable (Zanettin 2002). The following table summarizes these differences:

Table 1. Titles translated in Italy (from English) and in the United States (from Italian)

UDC category	E → I (Italy 76–95)		I → E (USA 77–96)	
	Texts	%	Texts	%
Literature/Children's Literature	4817	40%	502	28%
Art/Games/Sport	757	6%	343	19%
Education/Law/Social Science	1251	11%	187	11%
Applied Science	1835	16%	138	8%
History/Geography/Biography	919	8%	171	10%
Natural and Exact Science	643	6%	111	6%
Philosophy/Psychology	833	7%	53	3%
Religion/Theology	477	4%	267	15%
Generalities/Information Science	101	1%	2	0%
Total	11 633	100%	1774	100%

This table does not only indicate that translation from English into Italian is much more frequent than from Italian into English. It also shows that Italian non-fiction texts translated for the American market mainly belong to the domains of art/games/sports and religion, whereas American non-fiction texts translated into Italian have lower proportions from these domains, and a higher one of applied science texts.[2]

Now if we want our corpus to represent the operation of two different sets of translation policies (an arguably desirable objective), we need to follow the proportions set out above in each case, with the consequence that the various components will not be comparable. Alternatively, we can decide to make the corpus directional, and build it so as to represent the policies adopted in one direction only, or even not to bother about translation policies at all, and select texts opportunistically. Yet it would appear that in this way we miss the opportunity to relate the extra-textual conditioning factors of the context of situation/culture to the observation of linguistic patterning offered by corpora.

Monolingual comparable corpora (MCC) may appear to be untouched by these problems, since they do without source texts altogether. On the contrary, MCCs as built and used so far have involuntarily tended to obscure these realities, allowing texts to be detached from the extra-textual constraints (preliminary norms) that result in certain texts, writers, or genres having larger markets for translation than others in a given place at a given time. The bi-directionality criterion in CEXI and similar corpora, on the other hand, forces the corpus builder to face these problems, and solve them in some way or other: in the case of CEXI, it was decided that a common bi-directional core would be built, and then supplemented with directional sets that could be added to

the core depending on the priorities of each study. This is a welcome feature since, as we shall suggest below, the consciousness-raising function of corpus design and building is crucial if one is to avoid gross over-generalisations and misconceptions in using the corpus.

3.2   Preliminary norms 2: fiction

A second set of observations relates to the more subtle issue of the operation of preliminary norms in the selection of fiction texts. If the criteria adopted for selection are only those suggested by Baker (1995, above) and generally used in monolingual comparable corpus design – i.e. domain, language variety and time span – the result is hardly a comparable corpus, as the following lists illustrate:

First 20 fiction titles randomly selected from the Index Translationum *database (from English into Italian)*

1. Asimov, *The best science fiction of I. Asimov*
2. Atwood, *The handmaid's tale*
3. Christie, *The underdog and other stories*
4. Collins, *Rock star*
5. Dick, *The three stigmata of Palmer Eldritch*
6. Garfield, *The paladin*
7. Garrett, *Too many magicians*
8. Greene, *Shades of Greene*
9. Heller, *Catch 22*
10. Jong, *Fear of flying*
11. Koontz, *Strangers*
12. Le Carré, *The secret pilgrim*
13. Le Carré, *Tinker, tailor, soldier, spy*
14. McCarthy, *How I grew*
15. Smith, *The angels weep*
16. Smith, *When the lion feeds*
17. Stone, *Blizzard*
18. Strieber & Kunetka, *Warday*
19. Suyin, *The enchantress*
20. Van Lustbader, *Zero*

First 20 fiction titles randomly selected from the Index Translationum *database (from Italian into English)*

1. Calvino, *Gli amori difficili*
2. Calvino, *Il barone rampante*
3. Calvino, *Le cosmicomiche*
4. Calvino, *Se una notte d'inverno un viaggiatore*
5. Calvino, *Ti con zero*
6. Camon, *Un altare per la madre*
7. De Céspedes, *Il rimorso*
8. Eco, *Il nome della rosa*
9. Garofalo, *Operaio di Dio*
10. Lajolo, *Il vizio assurdo*
11. Levi, *Cristo si è fermato a Eboli*
12. Moravia, *1934*
13. Moravia, *Il conformista*
14. Moravia, *La ciociara*
15. Moravia, *La vita interiore*
16. Orseri, *Tikva. La porta della speranza*
17. Pavese, *Il mestiere di vivere*
18. Sciascia, *Candido*
19. Sciascia, *Il giorno della civetta; Il contesto*
20. Soldati, *La sposa americana*

For those acquainted with Italian and English literature it is clear that the two sets are not comparable at all.[3] Whereas the majority of the English texts could be described as low-brow or popular literature, the Italian texts are almost all classics, canonical exemplars from the production of high-brow authors. This is not, we would suggest, an effect of the text-selection procedure (random sampling ensures that no human bias was inserted at this stage), but a distinguishing feature of translation policies in the two cultures.

Let us now consider a monolingual comparable corpus, namely the pioneering Translational English Corpus (TEC) developed at UMIST, in particular its Italian component:[4]

Italian component of the Translational English Corpus

– Banti, *Artemisia*
– Buzzati, *Restless nights*
– Buzzati, *The siren*
– Capriolo, *The woman watching*

- Savinio, *The childhood of Nivasio Dolcemare*
- Tabucchi, *Declares Pereira*
- Tarchetti, *Passion*
- Tarchetti, *Fantastic tales*
- Troiano, *Jerome*

The source texts of some of these translations are very old: in several cases more than a century elapsed between the creation of the ST and the creation of the TT. Others have a somewhat doubtful status, for instance Troiano's *Jerome*, a book about translation, written by the founder and managing director of a translation agency, and translated into six languages to celebrate the twentieth anniversary of that agency. Four out of the ten texts have been translated by Lawrence Venuti, whose involvement with translation studies and ideas on the role of the translator (see e.g. Venuti 1998) may be a further source of bias.

It would not be impossible to envisage a comparable corpus of original texts which matched TEC in these respects. However, even in this case we would not have solved all our problems, since we would have to ask what the effects of this design decision might be. If we correct this "translation bias", which has to do with socio-cultural "preliminary" norms concerning what gets translated and why, are we then not potentially altering the picture to better suit our expectations? By restricting the choice of texts to include only those that are comparable we may be obscuring important situational differences between translated and original texts.

Let us consider what happens when the non-translated component is added to the Translational English Corpus, giving birth to the English Comparable Corpus. This component (non-TEC) is taken from the BNC. As such, it is made up of texts originally written in English and published between 1960 and 1993 (probably between 1975 and 1993, see Burnard 1995 for details). TEC-It, however, contains texts originally written in Italian and published between 1866 and 1994, translated into English in the late 1980s and 1990s. The underlying assumption here is that the date of original production of a text in language A does not affect the comparability of its translation into language B with an original in language B. Whilst this assumption may well prove correct, we would suggest it seems intuitively questionable, and in need of empirical verification.

A second point we wish to raise relates to the prestige of the works included in the two corpora. As we have seen, the texts that normally get translated from Italian into English (and that are hence more likely to make their way into TEC and similar corpora) tend to be very prestigious, canonical works in

the source culture. The works included in the BNC, however, were chosen to represent the language from a perspective of reception as well as production, and therefore include many bestsellers and widely-circulating books. Since the descriptive information such as "perceived quality status" and "target audience size" provided with (some of) the texts is too subjective and unreliable as a basis for comparison (Burnard, written personal communication, 2002), it is likely that the two sets of texts in the English Comparable Corpus are not members of one and the same population, distinguishable only by way of reference to the translation process undergone by one of them.

Again, this would not in itself be an insurmountable problem if corpus users were aware of this inherent bias, and tried to factor it out when attempting to interpret data. One suspects, for instance, that this mismatch might be one of the causes of Olohan's (2001) finding that "the language of TEC may [...] be judged as more formal" than the corresponding non-translational component of ECC. By examining the relevant source texts, this possibility might be checked and refuted, making the ample evidence in favour of explicitation as a "universal" feature of translation provided by Olohan all the more convincing. And if one had access to another MCC in a different language, the observation of explicitation processes in another context of situation/culture, subject to different preliminary norms, might enable one to base generalisations relating to laws of translation on much firmer ground.

Any feature characterizing a corpus of translations may be the result not just of the process of translation, but of the genres of the texts and of the influence of the source language or languages. The importance of genre and target audience for non-fiction texts has been shown by Mauranen (2000, 2002), who compares translated and original Finnish non-fiction texts from the academic and popular domains. As regards the role of the source language, it is clear that

> [w]hen studying translation as a product entirely in the target language environment, we can only put forward suggestions regarding the possible causes that may have led to certain patterns. In order to find an explanation for our results, we would need to construct and analyse in parallel another corpus that would include the source texts of the translational component [...]. (Laviosa 1998b: 565)

Summing up, if the status of a corpus of translations needs to be assessed against a comparable corpus of originals in the same language, it also needs to be assessed against the status of its source text in relation to a comparable corpus of original source language texts. For instance, we can only claim

that the type/token ratio in a corpus of translations provides evidence of simplification if (1) it is lower than that of a corpus of comparable original texts in the target language, and (2) this difference is greater than that between the type/token ratio of its source texts and that of a control corpus in the source language. This implies access to a large quantity and variety of electronic texts, to be combined in different ways within comparable corpora of different compositions.

For these reasons, we have decided to set up CEXI as a parallel bi-directional corpus allowing different combinations of subcorpora, in which each component can be used as a control for the mirror one.

4.   Conclusion

We hope to have shown that corpus-based translation research does not only involve word counts and software development, even though these are important aspects of the methodology. The search for norms or universals of translation through large quantities of texts is certainly favoured by corpus linguistics techniques, but it seems important not to forget that research based on specific types of corpora can only give us a partial picture, depending on what those corpora stand for. Corpora are an invaluable resource for the study of conventions, norms, and patterns of behaviour in different target cultures. But designing a translational corpus implies researching the social context(s) in which translations are produced and interpreted, so as to provide a framework within which textual and linguistic features of translation can be evaluated.

Extending the interpretation of findings based on a few texts and text combinations to postulate universal features of translation is likely to be mis-leading and counter-productive for the discipline. We can probably subscribe to de Beaugrande's (n.d.) claim about language universals, and extend it to translation universals as well:

> To judge from past experience, 'universals' tend to be indirectly extrapolated from particular languages after all, especially English. The latter's dominance in linguistic theory can only be effectively transcended by much resolute work on large corpora in as many languages as possible, each treated on its own terms.

Notes

1. Schmied and Schäffler (1996) prefer to use the term *explicitness* rather than *explicitation*. On this issue cf. Chesterman, this volume.

2. Our observations are limited to the American context because comparable British data for the same time span were not available from the *Index*, and other countries of publication had been excluded from the CEXI sampling frame.

3. Informal interviews of native speakers of English/Italian acquainted with Italian/English culture confirmed without exception our intuitions.

4. http://www.ccl.umist.ac.uk/staff/mona/tec.html#fiction, consulted 28/04/03.

References

Baker, Mona (1995). Corpora in Translation Studies: An Overview and some Suggestions for Future Research. *Target, 7*(2), 223–243.

Baker, Mona (1996). Corpus-based Translation Studies: The Challenges that Lie Ahead. In H. L. Somers (Ed.), *Terminology, LSP and Translation* (pp. 175–186). Amsterdam/Philadelphia: John Benjamins.

Beaugrande, Robert-Alain de (n.d.) Descriptive Linguistics at the Millennium: Corpus Data as Authentic Language. Online: http://www.beaugrande.com/, consulted 28/09/03.

BNC – British National Corpus. Online: http://info.ox.ac.uk/bnc/, consulted 28/09/03.

Burnard, Lou (1995). *Users' Reference Guide for the British National Corpus (SGML version)*. Oxford: Oxford University Computing Services.

CEXI – English Italian Translational Corpus. Online: http://www.sitlec.unibo.it/cexi/, consulted 28/09/03.

Chemnitz English-German Translation Corpus. Online: http://www.tuchemnitz.de/phil/english/chairs/linguist/real/independent/transcorpus/index.htm, consulted 28/09/03.

COMPARA – Portuguese English Comparable Corpus. Online: http://www.linguateca.pt/COMPARA/Welcome.html, consulted 28/09/03.

ENPC – English Norwegian Parallel Corpus. Online: http://www.hf.uio.no/iba/prosjekt/, consulted 28/09/03.

Even-Zohar, Itamar (2000 [1978/1990]). The Position of Translated Literature within the Literary Polysystem. This version originally published in *Poetics Today, 11*(1), 45–51. Reprinted in L. Venuti (Ed.), *The Translation Studies Reader* (pp. 192–197). London: Routledge. Online: http://www.tau.ac.il/˜itamarez/ps/pos_trli.htm, consulted 28/09/03.

Firth, J. R. (1935). On Sociological Linguistics. Extracted from Firth, J. R.: *The Technique of Semantics – Transactions of the Royal Society*. Reprinted in D. Hymes (Ed.), *Language in Culture and Society* [1964] (pp. 66–70). New York: Harper International.

Firth, J. R. (1956). Descriptive Linguistics and the Study of English. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth 1952–1959* [1968] (pp. 96–113). London/Harlow: Longman.

Hymes, Dell (1972). On Communicative Competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). London: Penguin.

Index Translationum, 5th edition. UNESCO 1998.

Laviosa, Sara (1997). How Comparable can 'Comparable Corpora' be? *Target, 9*(2), 289–319.

Laviosa, Sara (1998a). The English Comparable Corpus: A Resource and a Methodology. In L. Bowker, M. Cronin, D. Kenny, & J. Pearson (Eds.), *Unity in Diversity? Current Trends in Translation Studies* (pp. 101–112). Manchester: St. Jerome.

Laviosa, Sara (1998b). Core Patterns of Lexical Use in a Comparable Corpus of English Lexical Prose. *Meta, 43*(4), 557–570.

Laviosa, Sara (1998c). Universals of Translation. In M. Baker (Ed.), *The Routledge Encyclopaedia of Translation Studies* (pp. 288–291). London/New York: Routledge.

Mauranen, Anna (2000). Strange Strings in Translated Language. In M. Olohan (Ed.), *Intercultural Faultlines* (pp. 119–142). Manchester: St Jerome.

Mauranen, Anna (2001). Issues of Data in the Search for Translation Universals. *EST Conference 2001 Copenhagen Formats.* Online: www.cbs.dk/departments/english/EST/formats.shtml, consulted 24/04/02.

Mauranen, Anna (2002). Where is Cultural Adaptation? *Intralinea, 5.* Online: http://www.intralinea.it/vol5/cult2k/mauranen.htm, consulted 28/09/03.

Olohan, Maeve (2001). Spelling out the Optionals in Translation: A Corpus Study. In P. Rayson, A. Wilson, T. McEnery, A. Hardie, & S. Khoja (Eds.), *Proceedings of the Corpus Linguistics 2001 conference, UCREL Technical Papers: 13* (pp. 423–432). Lancaster: Lancaster University.

Olohan, Maeve, & Baker, Mona (2000). Reporting *that* in Translated English. Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures, 1*(2), 141–158.

Schmied, Josef, & Schäffler, Hildegard (1996). Explicitness as a Universal Feature of Translation. In M. Ljung (Ed.), *Corpus-based Studies in English: Papers from the seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)* (pp. 21–36). Amsterdam/Atlanta: Rodopi.

Sperberg-McQueen, Michael, & Burnard, Lou (2001). *The XML Version of the TEI Guidelines.* Online: http://www.tei-c.org/P4X/, consulted 28/09/03.

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

TEC – Translational English Corpus. Online: http://www.ccl.umist.ac.uk/staff/mona/tec, consulted 28/04/03.

Vanderauwera, Ria (1985). *Dutch Novels Translated into English. The Transformation of a Minority Literature.* Amsterdam: Rodopi.

Venuti, Lawrence (1998). *The Scandals of Translation: Towards an Ethics of Difference.* London/New York: Routledge.

Zanettin, Federico (2000). Parallel Corpora in Translation Studies. Issues in Corpus Design and Analysis. In M. Olohan (Ed.), *Intercultural Faultlines* (pp. 105–118). Manchester: St Jerome.

Zanettin, Federico (2002). CEXI. Designing an English Italian Translational Corpus. In B. Kettemann & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis* (pp. 329–343). Amsterdam/Atlanta: Rodopi.

Large-scale tendencies in translated language

Corpora, universals and interference

Anna Mauranen
University of Tampere

In the quest for translation universals, the status of interference has remained unclear. First, it is often indistinguishable from transfer, which blurs the concept of source language or source text influence on translated text. Second, it has been posited as either contradicting universals (Baker 1993), or as a universal, or a major translation law in itself (Toury 1995). This paper tackles these issues in the light of corpus data from the Corpus of Translated Finnish (CTF). It also offers a methodological path forward to comparing the relative distance of different corpora from each other, which is crucial for testing hypotheses concerning universals of translated language. The method is used for comparing the overall amount of transfer-like features in corpora from individual source languages, as well as from a mixture of several source languages.

1. Introduction

Mona Baker's seminal paper (1993) on translation universals has stirred both controversy and research activity in translation studies. The basic issues concerning the nature, or even the very existence, of universals in translation remains controversial, but Baker's original paper and a number of others following it have inspired fascinating research into fundamental issues in translation studies. One research project on these lines has been my own (see, e.g. Eskola 2002; Jantunen 2001; Mauranen 1998a, 2000a, Tirkkonen-Condit 2000). In the course of this research, one of the points of departure has been Mona Baker's definition of translation universals, which runs like this:

> universal features of translation, that is features which typically occur in translated texts rather than original utterances and which are not the result of interference from specific linguistic systems. (Baker 1993: 243)

The aspect in this definition that has begun to raise queries, particularly since it has come up in my own empirical work as well as that of my students, is the status of interference. Baker's definition appears to exclude interference, but if we turn to an earlier classic of translation universals (or 'translation laws' as he calls them), Gideon Toury, we see that he in fact posits "the law of interference" as a fundamental law of translation (Toury 1995: 275):

> in translation, phenomena pertaining to the make-up of the source text tend to be transferred to the target text.

Interference is thus either seen as contradicting universality, as in Baker's definition, or alternatively a basic manifestation of universality, as in Toury's. This is an intriguing conflict: are we dealing with different senses of 'universal', different levels or kinds of universal, or different understandings of 'interference'? Or possibly all of these?

Toury's other proposed universal, the law of growing standardisation, has under different guises received plenty of attention in the literature, while interference has remained in the shadow, perhaps in part due to Baker's formulation. In this paper, I would like to discuss two things: First, what do we understand by interference and in which ways can it be related to universals, and second, can we extract evidence from corpora to study this (and if, so, in which ways)?

2. Interference and its manifestations

The classic definition of interference comes from Uriel Weinreich: "those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language" (Weinreich 1953: 1). In second language learning, this has been taken to imply that an individual's first language (L1) necessarily influences his or her second language (L2), and an enormous amount of research has been devoted to describing and explaining the ways in which L1 interferes with L2. In contrast, translation studies, although more rarely referring to Weinreich, seem to have adopted a reverse view: it is the source language (the L2, as it were) that influences the target language (usually the translator's L1). Recently, some L2 acquisition scholars (papers in Cook 2003a) have been inspired by the implications of the phrase 'either language' in Weinreich's definition, and have started looking into the ways in which second (or third, etc) languages influence the first. This brings L2 acquisition research closer to translation

studies, but raises new issues. One of them is whether it is desirable or indeed possible to try to erase interference from translations.

Given that translation is a language contact situation, we might expect cross-language influence. It has been fairly well established that languages in contact generally influence each other (see, e.g. Thomason 2001). For example, Ellis (1996) points out that cross-linguistic influence appears to be present even at high levels of bilingual ability, and Grosjean and Soares (1987) have argued that when bilinguals speak one of their languages, the other language is rarely totally deactivated, even in completely monolingual situations. It is thus reasonable to assume, even without conclusive evidence, that transfer occurs in translation because translation involves a contact between two languages and is a form of bilingual processing. At a lower level of abstraction, more specific hypotheses can be posited, for instance concerning the levels of language where it is most influential (is it likely to affect syntax more than lexis or the level of discourse), to what extent it is local and textual (i.e. text-specific) and to what extent is it systemic (i.e. residing in the characteristics of the two language systems)? So far, it seems that transfer has been found in lexical, syntactic, pragmatic and textual phenomena, and thus all levels of language appear to be influenced. However, anecdotal evidence goes around among literary translators that it is the syntactic level that the SL most easily slips through. On the other hand, an earlier study (Mauranen 1999a) on translating existential themes suggested that translators typically sacrifice ST word order in favour of maintaining informational focus and TT textual flow.

The notion of interference itself appears somewhat vague, as currently used in translation studies. It sometimes seems to refer to SL influence on translations wholesale, that is, be roughly synonymous with, 'transfer'. But occasionally it is distinguished from transfer (e.g. Toury 1995: 252), which is taken to be the positive face of interference, which then is perceived as negative. It appears that "positive" transfer or just plain 'transfer' is more acceptable than "negative" transfer or interference. In fact Toury says himself that positive transfer is virtually indistinguishable from normal target language. The question therefore arises whether there is any reason (apart from possible theoretical ones) to deal with positive transfer? In a normative sense, we might simply accept its manifestations as 'good translation'. I shall return to this below.

For theoretical purposes, if transfer and interference are supposed to manifest the same underlying process, we naturally need to demonstrate that they are similar, and in turn distinguishable from 'non-transfer' translation. If we fail to do this, the concept of (positive) transfer loses its significance and becomes simply coextensive with 'translation'.

A general assumption seems to be that transfer is a relation between texts, that is, it occurs as influence from one text to another (e.g. Toury 1995). Even if it may also acquire systemic characteristics, these presumably take place where Toury's law of growing standardisation would apply, that is, at the TL end. Thus, to put it in Toury's terms, we replace (ST) textemes with (TL) repertoremes, but not vice versa.

However, we might question this assumption and posit instead that the source text activates the source language processing system, which in turn affects the target text production, because both the SL and the TL systems are simultaneously activated in the brain. That is, there need not be a direct influence at the level of text, but an indirect one from ST to SL system to TL system to TT. Some evidence for such a possibility comes from instances where a TT item looks like a likely candidate for transfer from the ST, but in fact has no stimulus in the source. For example Mankkinen (1999) was looking for anglicisms in a Finnish translation, and had picked items like *ottaa aikansa* ('it takes its time'), although the typical Finnish verb would be *viedä*, not *ottaa* (the gloss would be 'take' again, corresponding to a different sense of *take*). On inspection it turned out that the equivalent expression was not there in the source text; in other words, the apparent anglicism had not in fact been triggered off by the ST. A plausible explanation would therefore be that the bilingual processing situation activates both language systems, and that the source language system influences processing in the target language. Linguistic influence is, then, a normal consequence of language contact, or, part of what Cook (2003b:2) calls 'multicompetence' in a bi- or multilingual individual. If we were able to show that translation is an exception to this, that would be highly unexpected but of course all the more interesting.

How could this be shown, then? In other words, what kind of evidence would be needed for supporting an assumption that translations manifest no significant traces of interference from the source language? The first way in which this would receive support is if comparable corpora of translated and untranslated texts were sufficiently similar to each other to warrant the interpretation that we are talking about a single universe of texts. In statistical terms, since corpora are always samples, the question is whether they could have been drawn from the same population. We do not have entirely reliable statistical measures of overall differences or similarity in corpora yet (although for instance Kilgarriff is developing means for doing this, see Kilgarriff 2001), and before we do, we cannot address the question directly on an empirical basis. But I shall be exploring one possibility for such comparison a little later on (see, Section 6 below).

If we could show, then, that translations and comparable originals could have been drawn from the same text population, i.e. that they are samples from the same textual universe, this would imply not only that there is no significant interference but also that there are no (other) linguistic features which would systematically distinguish translations from originals written in the same language. In this way, the evidence would be more than sufficient, in fact too powerful for interference alone, and the case would be overdetermined. If, on the other hand, translations differ from originals, we cannot conversely automatically infer that the cause is interference. There may be other reasons, and so the evidence would be necessary but not sufficient. In short, only if translations in overall comparison are indistinguishable from similar TL texts, can we be certain that transfer plays no role in them.

If translations are distinguishable on a large scale from non-translated texts (as the evidence hitherto strongly suggests), an interpretation of the significance of interference derives from pitting it against universals altogether, so that the argument runs something like "instead of an universal language-independent law, we have 'pair-wise interference', that is, interference which is specific to the language pair in question, and which explains the 'oddity' of translations vs. original target language texts". This hypothesis, which reflects Baker's (1993) concept of universals, despite its opposite stance, would seem to receive support if a given feature can be observed in both a source text and a target text, but deviate from that which is typical in the TL. The research solution might be to start from individual, attested occurrences of interference. This would also seem to rescue us from the problem of positive transfer: if the results of transfer are hardly discernible from normal target-language productions, how do we distinguish the two? Toury (1995: 252) suggests that "the interference inherent in them becomes evident only when a translation is confronted with its source". If the assumed ST feature actually turns out to be behind the translation, it would seem to support the interpretation that a given source text has caused the translation (or more accurately, the transfer in the translation).

Yet, although the reasoning is intuitively satisfactory, it resembles the earlier assumption in second language acquisition research that the major cause of difficulties is interference from the learner's mother tongue (known as the "contrastive hypothesis"). It followed that the best predictor of interference problems would be contrastive analysis. However, on closer inspection it turned out that contrastive analysis was not very successful in predicting learner errors; as Mitchell and Myles (1998:30) put it: "the majority of errors could not be traced to the L1, and also [. . .] areas where the L1 should have

prevented errors was not always error-free." Thus, unless we can show that we have correctly predicted unusual occurrences of a given feature on the basis of SL-TL comparison, we are not on very firm ground in claiming that the pair-wise SL interference has been supported.

If the purpose is to show that bilingual interference provides more a powerful explanation of the linguistically special character of translations than more general, or 'universal' features of translation, which derive from the nature of the process and although possibly including interference are not limited to that, then fairly strong evidence is required to back it up.

First, we need to be able to correctly predict where interference occurs on the basis of a SL and a TL, and, moreover, where it does not occur. Such a prediction involves the systemic level of language; to maintain that interference obtains between a source and a target text, an analysis of the source text prior to seeing a target version of it should yield a text-specific prediction.

Second, we need to be able to test this on a parallel corpus. A parallel corpus, that is, one with source texts and their translations, is required to ascertain whether a particular target text feature, which we suspect of resulting from interference, actually regularly follows from a given, predicted, ST feature, and moreover, does not occur without this stimulus. If this is the case, we may be satisfied that its occurrence is connected with its source text, because such a finding indicates whether the feature is local – i.e. a consequence of the source stimulus.

Third, we need to be able to show that the resulting usage is exceptional with respect to the target language and translations from other source languages, as already pointed out above. If these three conditions are satisfied, it is warranted to say that the feature indeed occurs as a consequence of the ST stimulus, either only as a response to that, or at least more frequently than could be expected on the basis of normal TL practice alone, or even translated language more generally. This may look demanding but it hardly makes sense to grant the status of explanation on shakier grounds. Being able to successfully predict interference either at the level of language systems or at the level of individual texts would provide powerful evidence in favour of bilingual interference.

If, on the other hand, the purpose is not to show that bilingual interference overrides any law-like or universal tendencies, but rather to explore the plausibility of a general tendency towards transfer from a source to the target, it is not necessary to predict where exactly interference might occur; in fact this would be impossible, since the comparison would involve multiple source languages. The claim in such a case would be weaker in specificity, but stronger in generality; large-scale evidence which is compatible with interference as a

general tendency is fundamental to determining the status of interference in translation.

The data for showing this must primarily consist of comparable corpora, that is, matching corpora that have been compiled on the same principles of translated texts and original texts in the target language. Ideally, the data should comprise three kinds of (sub)corpora. First, a corpus of translated texts from one SL to one TL, to find out how frequent the postulated interference features are; secondly, a similar corpus with translations from different source languages, to ascertain whether the features in question are equally common, or more or less common, when translations come from various sources. Finally and most importantly, a corpus with comparable target language original texts is required to see whether the particular features are more common, less common, or equally common in TL original texts – in other words to check whether the occurrence of the features is exceptional on a large scale.

In this way it is possible to see whether there is anything remarkable in a potential 'interference' feature, that is, whether it occurs more frequently than could be expected on the basis of normal TL practice alone, or even translated language more generally. Such a comparison enables us to ascertain that there is something to explain (i.e. a deviation). At the same time, a more holistic view is maintained than by starting from ST–TT comparisons, and individual instances do not usurp an overblown importance. This procedure, then, allows us to test the assumption that systematic SL bias occurs in translation, which may then deserve the label interference (or transfer). Basically it allows us to assume that transfer /interference is likely to occur at the level of the language system, but it cannot show anything about the text-specific relations obtaining between particular STs and TTs.

3.   Interference or transfer – is there a difference?

As already pointed out above, 'transfer' and 'interference' are sometimes used interchangeably, sometimes as polar opposites. Interference in the latter case is seen as negative transfer, while transfer itself is held to be positive, or at least neutral. The distinction appears fuzzy, even arbitrary: if we have difficulty telling the positive from non-transfer, how do we distinguish positive from negative?

It seems to me that positive and negative transfer, insofar as both can be identified at various levels of linguistic description, can reasonably be conceived as points on a cline, one end of which is a gross deviation from

the target language norm, or what could be called a translation error in one sense, and the other is text which is indistinguishable in a normal reading from an original target language text, but in principle can be traced back to transfer from the ST, for instance through large-scale frequency differences (see, e.g. Gellerstam 1996; Laviosa-Braithwaite 1996). For normative purposes, the cline needs to be broken up somewhere, and a line drawn at some point where acceptable transfer is distinguished from unacceptable. Where to draw it is outside my present scope, but inevitably the question arises: perceived as negative by whom? Is it only normative translation specialists who determine what is negative and what is positive transfer as so often seems to be the case in the literature?

A more principled solution comes from Toury, who suggests that the acceptability is determined by social acceptance in the culture. He specifies this as a subcomponent of his law of interference as follows:

> tolerance of interference – and hence the endurance of its manifestations – tend to increase when translation is carried out from 'major' or highly prestigious language/culture, especially if the target language/culture is 'minor', 'weak', in any other sense[.]   (Toury 1995: 278)

On the basis of this, we would expect a difference for example in Finnish translations between English and Russian source languages. Presumably, and I think this is undeniable in present-day Finland, and has been for at least the decade that our corpus covers (the Corpus of Translated Finnish, CTF, see Section 4 below), that the English-speaking, Anglo-American culture is more dominant and generally more highly valued than the Russian culture. Therefore, if Toury's suggestion is right, Russian SL translations should deviate less from original Finnish than English SL translations because there should be a greater tolerance in the culture for English than Russian interference.

For testing this hypothesis, as well as the status of interference in relation to universality, I turn to the Corpus of Translated Finnish.

4.   The Corpus of Translated Finnish

The Corpus of Translated Finnish (CTF) was compiled at the Savonlinna School of Translation Studies 1997–2000 in my research group (Mauranen 1998a). It consists of 10 million words in all, about 4 million of which are texts of original Finnish and the rest translations from different source languages. The main source languages are English and Russian, and most of

the others (10 in all) are represented as one of two exemplars only. It is an important corpus for the study of translation universals, because it is one of the largest comparable corpora hitherto in existence, along with the pioneering Translational English Corpus (TEC) at the UMIST. It is also of special interest because the main language, Finnish, is not an Indo-European language. Most translation studies and most corpora that are currently available are heavily biased towards Indo-European languages, English in particular. Moreover, the CTF has been compiled as a comparable corpus from the beginning, therefore its compilation principles have remained consistent – to the extent that real-world conditions allow.

The amount of translations in the corpus is larger than that of originals because texts from different source languages can be compared to the same set of Finnish texts. The source languages are mostly Indo-European, but include Hungarian and Estonian, which, like Finnish, belong to the Finno-Ugric family. Issues of typological influence can therefore be studied with this data. The corpus contains whole texts, not extracts. The selection criteria were genre-based, or perhaps more precisely, domain-based, since the genres chosen are more appropriately described as genre clusters rather than basic-level genres (see, for example Mauranen 1998b), and the criteria remained external all through. Resorting to external criteria implied making use of publishers' and libraries' classification systems. This means fundamentally relying on a classification that is prevalent and generally accepted in the culture; we could also call this a set of 'folk genres'. Internal, or linguistic, criteria were deliberately avoided, because this would have meant selecting the data by the same criteria that would be used in its investigation. This would bring along serious problems of circularity, and although the folk genre approach may seem somewhat rough, it does reflect culturally relevant objects and meanings.

To ensure authenticity, the translations were all published, not elicited for the corpus. It was felt that high quality translations would be the first priority, since it makes sense to study translation products in a form in which they are accepted in the culture; therefore the texts came from established publishers, and the translators were mostly professionals. With some genres like academic texts, translators are usually experts in the field rather than professional translators, so an exception had to be made here. Since translation ideologies, traditions and fashions change, it was decided to opt for a narrow time window of five years (1995–1999), even though minor adjustments had to be made. The genres were chosen chiefly on account of their importance to translation. Three kinds of importance were distinguished:

1. that which is culturally influential on account of its *prestige* value (e.g. literary texts),
2. that which is influential by being *widely consumed* (e.g. popular and entertaining genres),
3. that which is influential by getting *translated to a large extent* (e.g. children's books, user's manuals and technical texts).

In practice the criteria were weighted also according to the interests of the research team. In consequence, the seven genres finally included are literary texts, academic texts, children's fiction, popular fiction, popular non-fiction, entertainment and biography. In practice the compilation was made with translations at the centre, since it was easier to find comparable Finnish to match these than the other way around. In all, these are what might be called 'prototypical' translations, which makes a good basis for a new field at the stage of exploratory studies.

5.   Comparing the corpora

To test the two assumptions discussed above, that transfer may be universal and that it is more acceptable if the source culture is dominant and has high prestige in the target culture, I chose to look at subcorpora of literature, because they offer the widest SL selection of matching texts:

The Subcorpora

Original Finnish Fiction
1 million

Translations with Mixed Source Languages (10 languages)
1 million +

Translations with English as Source Language
1 million +

Translations with Russian as Source Language
0.6 million +

The corpora are of a fairly even size, apart from the Russian subcorpus, which is clearly smaller, and just slightly over half the size of the others. This is a matter extremely hard to change – the time span for Russian was already extended backwards from the others (to the beginning of 1990s) to be able to

gather more data. Translation into Finnish is very much dominated by English sources – in fact the literary genre is an exception, having by far the greatest variety of SLs.

As pointed out above, there is no reliable measure of overall similarity and difference between corpora. I therefore developed a tentative solution for comparing the four subcorpora to one another, based on comparing lexis on a rank order basis. The point of departure was a frequency-based wordlist of each corpus. The lists consisted of individual word forms, not lemmata. Lemmatisation seemed unnecessary, even pointless, since it is by no means clear that great differences in the frequencies of typical forms are trivial from the point of view of translation – quite the contrary, in a highly inflectional language it can point to an untypical usage pattern (see, for example, Mauranen 1999b).

Since the method of comparison needed for this study had no precedent to go on, I started from some general principles that we already know from corpus linguistics (for alternative solutions, see Jantunen, this volume). First, the most frequent items in different reasonably-sized corpora tend to be fairly consistently the same in a given language. In fact, frequent items can show remarkable consistency even across highly unequal corpus sizes, up to a 100-fold difference (see, e.g. Mauranen 1999b). Therefore, the fact that the Russian SL subcorpus was only about two thirds the size of the others, should not influence the results dramatically, especially as rank order is not very sensitive to size. Second, fairly soon after the top frequency words, corpora begin to show their differences, and the high frequencies peter out and tail off into a very long list of few, finally single occurrences.

My solution was to try out three different frequency bands of thirty words each; the first, from 1 to 30, the second from 50 to 79, and the third from 100 to 129 (with some adjustment on account of excluding proper names). The bands were chosen conservatively in that they were all fairly high in frequency: this meant it was possible that there would not be much variation. My assumption was that given the distributions of monolingual corpora, there would be very little variation in the top frequency band, somewhat more in the second, but the last one which was picked from just below the 100 top frequency level, would be the least predictable. I also assumed that the best guess for finding meaningful variation would be in the middle band. Of course, although we might reasonably expect this to be the case, the exact optimal place for the middle band is ultimately an empirical matter, and no precedents existed to look into.

After setting up the frequency bands, and excluding the proper names, the procedure was as follows. At the first stage, the original Finnish texts' rank ordered vocabulary was adopted as the standard for comparison, i.e. the reference corpus. The other corpora were compared to this by noting the deviation of each item from the standard, that is, the difference in the item's rank order position from the position of the same item in the standard. The deviations occurred either upwards (the item had a higher rank in the translation corpus than the standard), or downwards (the item's rank was lower), which meant that they might have cancelled each other out unless only their eigenvalues (points of difference in the rank) were noted. The eigenvalues were then summed for an aggregate estimate of the difference between the reference corpus and each subcorpus at a time. The same procedure was then applied to comparisons between translational corpora at the second stage, with the subcorpus of mixed languages as the reference corpus.

This experimental method was developed as a tentative measure for comparing the relative distances between corpora. It goes without saying that such a measure remains partial because it is based on lexical rank order differences only, but in the absence of comprehensive overall measures it can be used as a pointer, in conjunction with other measures where possible. On account of its exploratory character it is not well suited for existing tests of statistical significance – even nonparametric tests which in principle might be considered, make such assumptions about the populations which do not apply to data consisting of a mass of running text. What we can hope from the present method is, then, a rough outline of the degree to which corpora might differ from each other, and expect the outline to be filled out with complementary means.

6.   Findings

Applying the comparative method above to the subcorpora resulted in some interesting, even surprising observations. What we find is that it was the medium-frequency band (50–79), not the lowest, which actually shows the greatest overall differences (Tables 1–4 below). However, there was some variation. Thus, in comparison with original Finnish, the English subcorpus is an exception: it shows a steady increase of deviations as we go down the frequency list.

As a test of universality vs. SL specific interference, I suggested above that if universality overrides bilingual interference, there should be little difference between texts from mixed SLs and texts from particular SLs, but all of these

Table 1. Sum of differences from original Finnish

Freq. Band	Mixed	English	Russian	Σ
1–30	87	75	96	258
50–79	142	87	178	407
100–129	62	167	77	306
Σ	291	329	351	971

Table 2. Sum of differences from mixed source languages

Freq. Band	English	Russian	Σ
1–30	63	71	134
50–79	190	115	305
100–129	104	51	155
Σ	357	237	594

should be distinct from original Finnish. In other words, the three translated subcorpora should be more similar to each other than to the original Finnish corpus. Table 1 shows that the basic assumption of all three corpora deviating from originals is supported. This is hardly a surprise. What is more interesting is that there are also individual patterns: Mixed SLs deviate the least, Russian the most, and English is in the middle.

Let us now see what happens if we compare the individual SLs to the mixed-language translation corpus (Table 2). Here is a clear difference between the two: Russian appears to be closer to general translationese than English.

The less predictable question that we asked above was whether the individual SL subcorpora deviate more from the originals than they deviate from translations on the whole. If interference from particular SLs is a more influential factor than translationese on the whole, the differences between the various translational sources ought to be greater than those between translations and originals. In Table 3, I have combined figures from Tables 1 and 2, comparing the English and Russian subcorpora to mixed SL translations on the one hand and to originals on the other.

The overall figure for deviations from the reference corpora is indeed clearly higher for originals than for translations. That is, the translations from individual SLs are more like translations on the whole than they are like original Finnish. This provides support for the hypothesis that translations share features that distinguish them from original texts in the same language. Thus, the present findings suggest that translations show a certain affinity to each other; it follows that 'translationese', or the deviation of translations from

Table 3. English and Russian Translations compared to mixed source languages and original Finnish

Freq. Band	Mixed source languages			Originals		
	English	Russian	Σ	English	Russian	Σ
1–30	63	71	134	75	96	171
50–79	190	115	305	87	178	265
100–129	104	51	155	167	77	244
Σ	357	237	**594**	329	351	**680**

Table 4. English and Russian Translations compared to original Finnish

Freq. Band	Originals		
	English	Russian	Σ
1–30	75	96	171
50–79	87	178	265
100–129	167	77	244
Σ	329	351	680

TL originals cannot be reduced to SL-specific interference. At the same time, there is a clear profile difference between the source languages: while English SL texts deviate less from Finnish originals than from other translations, translations from Russian show the reverse tendency. This suggests traces of SL-specific interference. Thus, the results are compatible with the interpretation that interference is universal. In sum, the present findings suggest that overall, translations resemble each other more than original target language texts, but a clear source language effect is also discernible. This implies that transfer is one of the causes behind the special features of translated language.

Finally, what about the differences between translations from English and from Russian? The hypothesis was that Russian SL texts should deviate less from original Finnish than English SL texts because there should be a greater tolerance in the culture for English than Russian interference. In fact, if we compare them (Table 4 is a repeat of the right-hand side of Table 3 above), we notice that Russian deviates more from original Finnish, not less. Thus the hypothesis of more deviation being accepted from a prestige culture receives no support from this data.

Obviously, there is the weakness that there is less data from Russian. This in itself of course shows that the prestige value of Russian is lower, but as things stand, this bias cannot be hoped to be corrected; it is probably endemic. Getting equal amounts of data from more peripheral and more central source cultures

is likely to remain low. In statistical terms, however, the impact of unequal corpus size is much reduced by the fact that the comparisons are based on the rank order differences, not direct frequencies. The result is intriguing because it runs counter to Toury's perfectly reasonable assumption. It calls for further research and new explanations.

7.   Conclusion

It has been argued in this paper that in order to explore the plausibility of interference constituting a fundamental law of translation, or a translation universal, it is necessary to have access to different kinds of comparable corpora: original texts in the target language, and translations with different source languages. The findings based on such comparable corpora indicated that translated texts deviated clearly from the original, untranslated texts, and on the whole, translations bore a closer affinity to each other than to untranslated texts. At the same time, different source languages, Russian and English, showed individual profiles of deviation. The results suggest that the source language is influential in shaping translations, but it cannot be the sole cause, because the translations resembled each other. The study therefore lends support to Toury's (1995) claim that interference or transfer constitutes a general law of translation. It also supports Baker's (1993) hypothesis insofar as the bilingual interference between particular language pairs does not seem to exhaust the differential between translations and non-translations. To reconcile the two hypotheses we simply need to recognise that the general tendency of source language influence on translations is an abstraction based on a number of language pairs showing the same trend; whereas the influence of a particular source language (or indeed source text, as is also often assumed) on a particular target language is not sufficient to account for the differences between translated and untranslated texts. Therefore, interference (or transfer) is best conceptualised as one of the universal tendencies, on a high level of abstraction, precisely on account of predictably taking place in each language pair involved in translation.

The general-level comparison carried out in this study cannot pinpoint individual occurrences of interference. Intriguing research questions therefore remain: is transfer universal because it involves bilingual processing and therefore an inescapable contact between two language systems, a consequence of the 'multicompetence' (Cook 2003b) of a multilingual individual? Or is it triggered off by the source text, and the translator's task of rendering that text in a

new guise? More precise understanding of whether different levels of language are affected differently by interference is also needed. It seems, for example, that pragmatic interference can exert a strong influence on target texts (e.g. Mauranen 2000b).

Do we need to make a systematic distinction between (positive) transfer and (negative) interference? In this paper the terms have been used interchangeably, but it has been suggested (Eskola 2002, this volume) that we might redefine interference as a neutral, descriptive term. But since a non-negative term already exists, would it not be preferable to continue using that? One possibility of distinguishing the two might be to employ transfer to refer to the exaggeration or overrepresentation of shared features between the SL and the TL, i.e. 'preferred choices', or unmarked choices in both. Interference would then be reserved for deviation from TL norms towards the SL norm, i.e. 'dispreferred features' in the TL. Examples of the latter would be collocations or other combinations which break no obvious rule of the TL but are simply not found in original texts (see, e.g. Mauranen 2000a). The distinction would hardly become entirely clearcut, but one distinct advantage would be a clearer formulation of hypotheses that have a bearing on universal tendencies, such as for example the one discussed in this paper. To make further progress towards capturing universals, we might then want to hypothesise that transfer phenomena are more widespread than interference phenomena. This would imply that features shared by the source and the target languages would have a proportionally stronger representation in translated texts than originals, while the same would not be true of features where the two languages differ.

The test for cultural dominance affecting acceptability failed to produce the expected outcome. A number of alternative explanations spring to mind: Finnish may already be influenced by English, therefore the smaller distance; or established older translation traditions from Russian may influence present practices. To begin to find answers, we need to delve deep into social and historical contexts of translation, possibly into historical translation corpora.

References

Baker, Mona (1993). Corpus Linguistics and Translation Studies – Implications and Applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology. In Honour of John Sinclair* (pp. 233–250). Amsterdam: John Benjamins.

Cook, Vivian (2003a). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters

Cook, Vivian (2003b). Introduction: The Changing L1 in the L2 User's Mind. In V. Cook (Ed.), *Effects of the Second Language on the First* (pp. 1–18). Clevedon: Multilingual Matters.

Ellis, Nick C. (1996). Sequencing in SLA. Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition, 18*, 91–126.

Eskola, Sari (2002). *Syntetisoivat rakenteet käännössuomessa: suomennetun kauno-kirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä* [Synthesising Structures in Translated Finnish. A Corpus-based Analysis of the Special Features of Finnish Literary Translations], [Publications of the University of Joensuu 30]. Joensuu, University of Joensuu.

Eskola, Sari (this volume). Untypical Frequencies in Translated Language: a Corpus-Based Study on a Literary Corpus of Translated and Non-Translated Finnish. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Gellerstam, Martin (1996). Translation as a source for cross-linguistic studies. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast* (pp. 53–62). Lund: Lund University Press,

Grosjean, Francois, & Soares, Carlos (1987). Processing mixed language: some preliminary findings. In J. Vaid (Ed.), *Language Processing in Bilinguals: Psycholinguistic and Neurophysiological Perspectives* (pp. 145–179). Hillsdale, NJ: Erlbaum.

Jantunen, Jarmo (2001). Synonymity and lexical simplification in translations: A corpus-based approach. *Across Languages and Cultures, 2*(1), 97–112.

Jantunen, Jarmo (this volume). Untypical Patterns in Translations: Issues on Corpus Methodology and Synonymity. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Kilgarriff, Adam (2001). Comparing corpora. Online: http://www/itri.brighton.ac.uk, consulted August 26, 2001.

Laviosa-Braithwaite, Sara (1996). *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*, UMIST: Unpublished PhD thesis.

Mankkinen, Marika (1999). *Manifestations of translatedness: a case study of anglicisms in the novel Trainspotting*. Unpublished MA Thesis, Savonlinna School of Translation Studies, University of Joensuu.

Mauranen, Anna (1998a). *Käännössuomi ja kääntämisen universaalit: tutkimus korpusai-neistolla* [Translated Finnish and Translation Universals: A Corpus Study], Project Research Plan, MS, Savonlinna School of Translation Studies.

Mauranen, Anna (1998b). Another Look at Genre: Corpus linguistics vs. genre analysis. *Studia Anglica Posnaniensia, XXXIII*, 303–315.

Mauranen, Anna (1999a). What sort of theme is *there*? A translational perspective. *Languages in Contrast, 2*(1), 57–86.

Mauranen, Anna (1999b). Will 'translationese' ruin a contrastive study? *Languages in Contrast, 2*(2), 161–185.

Mauranen, Anna (2000a). Strange Strings in Translated Language: A Study on Corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects* (pp. 119–141). Manchester: St. Jerome Publishing.

Mauranen, Anna (2000b). Where's cultural adaptation? Paper presented at CULT 2K, the 2nd International conference on Corpus Use and Learning to Translate. University of Bologna 3–4.11. 2000, Bertinoro.

Mitchell, Rosamond, & Myles, Florence (1998). *Second language learning theories.* London: Arnold.

Thomason, Sara Grey (2001). *Language Contact.* Edinburgh: Edinburgh University Press.

Tirkkonen-Condit, Sonja (2000). In search of translation universals: Non-equivalence or 'unique' items in a corpus test. Paper presented at Research Models in Translation Studies. UMIST 28–30 April, 2000, Manchester.

Toury, Gideon (1995). *Descriptive Translation Studies – and Beyond.* Amsterdam: John Benjamins.

Weinreich, Uriel (1953). *Languages in Contact.* The Hague: Mouton.

Untypical frequencies in translated language

A corpus-based study on a literary corpus of translated and non-translated Finnish

Sari Eskola

University of Joensuu

The theoretical goal of this paper is to clarify some central concepts frequently used in corpus-based translation studies. When we are primarily interested in uncovering the essence of translation *per se*, we should not make a distinction between norm-dependent and potential universal features but rather talk about laws of translation more widely (as both local and global inherent tendencies and regularities pertaining to translation). The empirical goal is to outline some results concerning dissimilarities in the frequencies and distributions of three non-finite structures of the Finnish language (referative, final and temporal constructions) in different language variants: texts originally produced in Finnish and texts translated from English and Russian into Finnish. I provide evidence in support of a possible universal law that translations tend to under-represent target-language-specific, unique linguistic features and over-represent features that have straightforward translation equivalents (functioning as some kind of stimuli) in the source language. It is a question of interference but not in a negative, but rather a neutral, abstract and statistical sense.

1. Introduction

There has been a gradual shift from prescriptiveness in translation studies towards understanding that translations inevitably form a language variant of their own: they tend (and are also allowed) to possess properties that differ from those of texts that have originally been produced in the same language (translations are "different", not "deviant" as Baker 1999: 292 puts it). Translated texts have been referred to as "the third code" (Frawley 1984), "the third language" (Duff 1981) and "hybrid language" (Trosborg 2000). However,

our knowledge of concrete distinctive features of translations is still vague, and the question remains what really makes them the way they are. Before detailed statements can be made on the subject, we definitely need more profound and systematic comparative research on translated and non-translated texts based on large electronic text collections and corpus methodology.

Not long ago Baker (1993, 1995, 1996) launched the idea of using the methods of corpus linguistics in order to uncover the distinguishing features of translated language. Now a growing number of researchers work in the new field of corpus-based translation studies (CTS), trying to capture the real nature of translated texts and bring something concrete to the rather obscure discussion conducted (critically or otherwise) in the literature on translation and also among the general public. This kind of descriptive study is greatly facilitated by the availability of corpus linguistic tools. Different corpora allow one to analyse language in a real context on both a quantitative and a qualitative basis, and the application of corpus linguistics can reveal something about translations that we have not been able to see using small corpora and manual methods.

2.   From norms to laws

Much effort has been devoted to the vexed question of *norms* (e.g. Toury 1978, 1980, 1985; Schäffner 1999; Chesterman 1997), and not least in CTS. The concept itself has been adopted from social sciences to translation studies and there is still no agreement in the literature as to what exactly constitutes norms of translation. One of the main problems seems to be that norms are often equated with observed regularity, which is why too many things are considered norms and the concept itself has suffered and lost its explanatory power. In my view, norms are not themselves observable but can be identified on the basis of regularities in recurrent situations. The very essence of norms is that they are binding constraints, social expectations that tell us how to behave and against the backdrop of which our behaviour can be evaluated. Norms result in regularities of behaviour, but linguistic features themselves are not norms. Even if norms can be identified on the basis of regularities, regularity itself is not necessarily a proof of the existence of a norm, because it may also have other causes. Identifying what features actually are norm-dependent requires that we find links between knowledge of values and priorities on the one hand and features that are observable in translations on the other (see Pym 1998).

In CTS the concept of norm has, alongside that of universals, become central as one explanation for repeated patterns found in translations. Many commentators refer to local and conditioned regularities of behaviour as norms or norm-dependent phenomena (e.g. Kohn 1996; Baker 1993; Øverås 1998); while norms operate in local socio-cultural contexts and change over time, universals are globally observable tendencies and regularities of behaviour that can be found in translations irrespective of the languages involved. With respect to some features of translation there seems to be confusion about whether they are norm-dependent or universal (for example explicitation, see Vanderauwera 1985; Blum-Kulka 1986; Weissbord 1992; Øverås 1998). In my view, norms are primarily prescriptive by nature, while universals are descriptive and predictive, and this is why we should not use these terms as alternative explanations for regular distinguishing features of translations, and by doing so restrict the potential of CTS unnecessarily.

It would be really important, then, to start to talk about translation *laws* more widely (a very good concept put forward initially by Toury 1991 but rather little used in translation studies in general): if we want to find out how translations *per se* deviate from texts that have been originally written in the target language and how translation as a specific process influences linguistic behaviour, the main object of interest also locally, under particular conditions, is *not* norms but rather laws of translation, features that are inherent in translation. Consequently, I would rather make a distinction between local and universal translation laws than talk about norms and universals as parallel phenomena. Local laws can be found for example in a certain language pair, text type and time span, whereas universal laws are global tendencies that operate in all translation. The impact of the translation process may result in statistical preferences and characteristics that are distinctive of translating between languages A and B for instance. Rabin (1958: 144–145) argues that translators of a certain language pair may build up a kind of "translation stock" of tried and tested strategies and this can subsequently mark such translations. Behind such local features, there might be some universal tendencies that operate in all translation. On similar lines Chesterman (1998) speaks about laws that indicate what either all translators in general or some subset of them tend to do. He also states that "the task of empirical research is then to establish the conditions under which such laws seem to hold, and with what probability, or under which they do not hold" (ibid. 218).

Corpus linguistic techniques can bring out observable regular patterns in translations, and on that basis one might also want to speculate about which norms may have influenced the features that are found. As norms

have an impact on translators' choices in actual situations, they also influence translation laws. In a methodological sense, then, norms have explanatory force, but it is always up to the researcher to *interpret* what norms have been applied. Norms also can be universal by nature (in contrast to individual or otherwise local norms), but we should not confuse them with inherent universal tendencies (laws).

It must be something in the nature and process of translation that causes translation laws. This 'something' is still quite vague in translation studies as we do not know exactly what it might be. I do not mean here the formalized models and theories of the translation process designed to describe how translators progress in their work, but rather the basic difference of the nature of translation as a cognitive process in contrast to original writing. According to Klaudy (1995: 142) the road leading from the mind to the linguistic form is never direct and simple even if we operate in our mother tongues; if the thought takes its origin in another language the linguistic process is inevitably more complex and bound by a larger number of constraints. Translation, then, is a complex transaction and there are several factors that have an impact on it: at least, distinctive features of ST, SL and TL, the translation tradition (including norms) and also individual preferences. These are all local features. The more global and abstract the law, the clearer the impact, so to speak, of the nature of translation as a unique linguistic process as such and the smaller the possible impact of the source language, text type etc. In other words the impact of these above-mentioned factors on the process is more obvious in local than in universal laws. Universal laws (e.g. of simplification, explicitation and conventionality) are not necessarily absolute laws, but strong statistical tendencies that can be observed widely (showing what translators on the average tend to do and what they do not tend to do). So far they have been mostly identified intuitively and by small-scale, manual analyses and need to be examined critically. Hypotheses about universals can be verified only if we get results on the basis of several language pairs (preferably also other languages than Indo-European) and different kinds of linguistic elements (lexical, syntactic, textual, stylistic). Studying translation universals is like trying to solve a jigsaw puzzle. Every piece of information about the use of any single pattern is part of the whole when we try to find out what translations are really like. In addition, every individual study also provides valuable information about a specific text type and language pair, and about typicalities that operate in translation at the local level.

Table 1. The Finnish Corpus of Translational and Non-translational Narrative Prose

Narrative prose	No of Texts	No of Words
Translations from English (TE-texts)	11	639,608
Translations from Russian (TR-texts)	11	635,511
Original Finnish (OF-texts)	19	619,296
Total	41	1,933,279

3. Use of referative, temporal and final constructions in translated and non-translated texts

3.1 Data

As Table 1 shows, my corpus (*The Finnish Corpus of Translational and Non-translational Narrative Prose*) consists of three different components (language variants): original Finnish narrative prose (*OF-texts*) and narrative prose translated from English (*TE-texts*) and Russian (*TR-texts*) into Finnish. The data are subcorpora of the *Corpus of Translated Finnish* (*CTF*) compiled at the Savonlinna School of Translation Studies. All of the texts have been published in the 1990's and they are full, unabridged texts, not text fragments. The size of the corpus is about 2 million words and the word-count of each of the components is approximately 600,000 (since each component sample is equal in size, the results are directly comparable).

In Finland translations form a substantial part of written texts and translations are widely read. Approximately 60% of all published narrative prose in Finland is translated and there is a huge difference between English and Russian as source languages in this respect. About 70% of all translations are translated from English and only 1% is translated from Russian. Therefore I have in my corpus source languages that have quite different translation traditions in Finland: there are differences in the way they are (and are expected to be) translated and thus the norms operating in these translation traditions deviate from each other.

3.2 Results

In my doctoral dissertation (Eskola 2002) I compare translated Finnish language with original texts, trying to examine both local and global translation

laws. In this sense my research progresses from concrete to abstract and from local to global. On the local level I draw conclusions about regularities of translators' behaviour in given language pairs (English-Finnish, Russian-Finnish), a given time span (contemporary literature) and text type (literary prose). The aim of using two different translational subcorpora is to examine the possible impact of the source language on translated Finnish. Results concerning features that are found irrespective of the source languages can be used to test hypotheses concerning universals of translation on the global level.

This paper concentrates on three non-finite syntactic structures, namely referative (e.g. *Tiedän hänen tulleen* 'I know she has come'), temporal (e.g. *Lukiessaan kirjaa* 'Reading a book') and final (e.g. *Kiirehdin ehtiäkseni junaan* 'I hurried to catch the train') constructions. These are packed predications which are often used to compress information. They do not include a finite verb and could alternatively be realized by a subordinate clause: the finite and non-finite variants cover the same information and are typically considered as interchangeable. As there is an option available in the use of these structures it is interesting to find out differences in patterns of choice in their use in translated and non-translated texts. Compared to many Indo-European languages, the Finnish language is very synthetic and uses structures of these kinds productively.

The starting point of my analysis is the hypothesis that translations tend to show untypical syntactic, lexical and textual frequencies as compared to non-translated texts. There are some results supporting this law but they are still quite few (e.g. Gellerstam 1996; Laviosa-Braithwaite 1996; Mauranen 2000). The results at hand focus on untypical frequencies on the syntactic level, and a central factor is the availability or absence of corresponding syntactic elements in the source language. There is a clear tendency that preferences in choosing between certain interchangeable expressions in translations are strongly associated with the features of the source language, both in terms of the systemic possibility and of the actual typicality of corresponding constructions. Whereas contrastive research on the typicality of particular linguistic structures in different languages is still largely missing and intuition is not a very good tool for estimating it, the knowledge of differences and similarities of the systemic features of languages is on a much firmer basis. In this sense the analysed Finnish non-finite verb forms can be divided into the following subgroups:

a. The structure is unique and language-specific; there is no straightforward equivalent in English and Russian that could be productively paraphrased by a finite verb form (referative construction).

b. Despite certain restrictions, the structure has an equivalent in English and Russian that can be productively paraphrased by a finite verb form (temporal construction).

c. The structure has a clear straightforward equivalent in English and Russian that has no productive finite alternative (final construction).

The factors mentioned are all common to English and Russian; Finnish is in this sense quite different from both. However, I will later mention features that differentiate English and Russian and show that it is often the dissimilarities between these source languages that cause differences between translations from them. Now I will present my results concerning the differences in frequencies and distributions of referative, temporal and final constructions in translated and non-translated texts in more detail.

3.2.1   *The referative construction*

The referative construction is used in Finnish to contract an affirmative *that*-clause with verbs such as *see, hear, believe* and *say* etc. It represents a syntactic structure which is specific to the Finnish language and which has no straightforward equivalent in Russian and English. As examples (1)–(2) illustrate, in Finnish you can choose between a finite verb form and its compact, non-finite counterpart (irrespective of the verb), but in Russian and English it is typical to prefer either a non-finite or a finite verb form in referative expressions, and a choice between interchangeable variants is quite rare (in English the verb *see* requires a non-finite and *know* a finite verb form, in Russian corresponding expressions call for finite verb forms in both cases).

(1) F a.   *Näin Liisan <u>lukevan</u> kirjaa.*   non-finite
    b.   *Näin, <u>että Liisa lukee kirja.</u>*   finite
  R   *Ja videla, čto Liisa čitaet knigu.*   finite
  E   *I saw Liisa <u>read/reading</u> a book.*   non-finite

(2) F a.   *Tiedän hänen <u>tulleen.</u>*   non-finite
    b.   *Tiedän, <u>että hän on tullut.</u>*   finite
  R   *Ja znaju, čto ona prišla.*   finite
  E   *I know (that) she has come.*   finite

The results show that translations have a lower frequency of referative constructions than the original Finnish texts (Figure 1). This tendency is especially strong in translations from Russian. The under-representation of referative constructions in translations seems quite logical as there is no systematic infinitive stimulus in corresponding structures in the source languages.
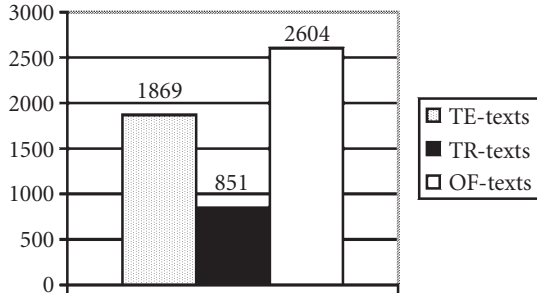
Figure 1. Frequencies of referative constructions

The time referred to in referative constructions shows some interesting differences between translations from different source languages (Figures 2–3). In translations from Russian the frequencies of referative constructions in both the present and past tense are lower than in original texts. In translations from English especially structures used in the past tense are clearly underrepresented, but in the present tense differences between translations from English and original texts are not so large.
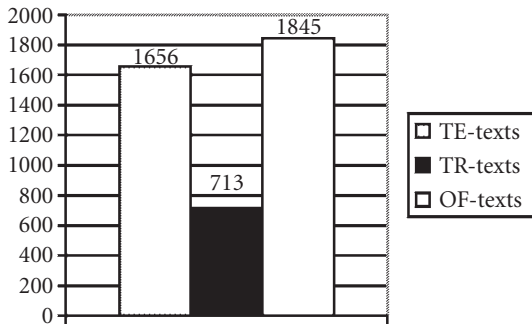

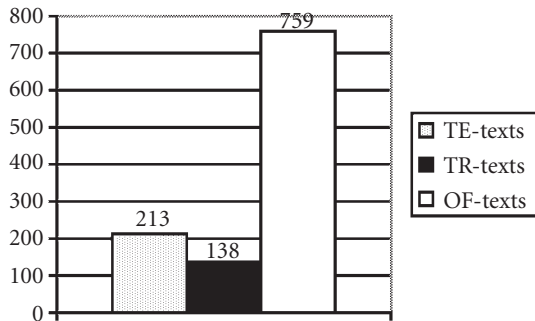**Figure 2.** Frequencies of referative constructions in the present tense

Figure 3. Frequencies of referative constructions in the past tense

In addition to showing that the referative construction is used less frequently in translations, the results indicate that it is used also in a different way. Tables 2 and 3 show the frequencies of two specific types of referative constructions that show some appreciable dissimilarities between translated and non-translated texts. First, there is a clear tendency that referative constructions which are used with perception verbs (e.g. *see, hear, notice*) in the present tense are strongly overrepresented in translations from English (Table 2): the figures show that both the frequency and the relative proportion of such structures are higher in translations from English than in the other two components.

Table 2. Frequencies and percentages of perception verbs in referative constuctions (present tense)

TE-texts N = 1869		TR-texts N = 851		OF-texts N = 2604	
f	%	f	%	f	%
521	27.9	123	14.5	236	9.1

This explains at least partly the differences between translations from Russian and English in the frequencies of the referative construction used in the present tense (shown in Figure 2). As opposed to Russian and Finnish, in English perception verbs are typically used in referative expressions with an infinitive (3). On the basis of these results this systemic feature seems to relate to their overuse in translations from English.

(3)   **Näin** Liisan <u>lukevan</u> kirjaa.   non-finite
      Ja **videla**, čto Liisa čitaet knigu.   finite
      I **saw** Liisa <u>read/reading</u> a book.   non-finite

Second, there is a clear tendency for referative constructions that are used with verbs of saying and reporting (e.g. *say*, *tell*, *inform*) in both tenses to be used less in both translated text groups than in original Finnish (Table 3).

Table 3. Frequencies and percentages of verbs of saying and reporting in referative constructions (present and past tense)

TE-texts N = 1869		TR-texts N = 851		OF-texts N = 2604	
f	%	f	%	f	%
181	9.7	80	9.4	1036	39.8

In Russian and English such verbs are typically used in referative expressions with finite verb forms (4), so it seems to be the case that there is no stimulus for this type of referative construction in either of the source languages. It is by far the most common type of referative construction in original Finnish texts and quite rare in translations.

(4) *Hän **kertoi** minulle <u>tulevansa</u>.*   non-finite
     *Ona **skazala** mne, čto pridet.*   finite
     *She **told** me that she was coming.*   finite

3.2.2   *The temporal construction*

The temporal construction is used to indicate the time of the action in relation to the main clause. In many cases the corresponding structures in Russian and English offer a choice between non-finite and finite structures both in the present and the past tense (5–6). There are, however, certain restrictions: in Russian and English such structures (gerunds and *ing*-participles) are not used when the subject of the non-finite expression is not the same as the subject of the main clause. In English there is, however, an expression called the 'absolute participle structure' by Zandvoort (1975: 35–36), which means a participle construction having a different subject than the main clause (e.g. *The authorities having arrived -.-, the ceremony began*).

(5) F a.   <u>*Lukiessaan*</u> *kirjaa…*   non-finite
     b.   *Kun hän lukee kirjaa…*   finite
     R a.   <u>*Čitaja*</u> *knigu…*   non-finite
     b.   *Kogda ona čitaet knigu…*   finite
     E a.   <u>*Reading*</u> *a book…*   non-finite
     b.   *As she is reading a book…*   finite

(6) F a.   <u>*Luettuaan*</u> *kirjan…*   non-finite
     b.   *Kun hän on lukenut kirjan…*   finite
     R a.   <u>*Pročitav*</u> *knigu…*   non-finite
     b.   *Posle togo kak ona pročitala knigu*   finite
     E a.   <u>*Having read*</u> *the book…*   non-finite
     b.   *When she has read the book…*   finite

The frequencies of the temporal construction show that it is used more in both translated components (Figure 4). Differences between translations from English and Russian are negligible.
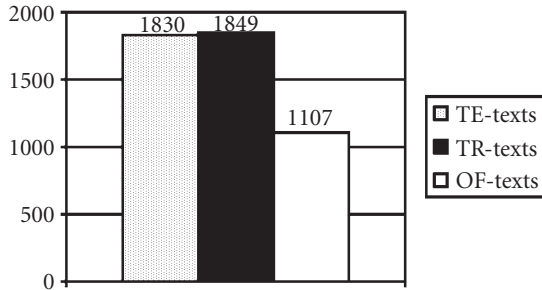
Figure 4. Frequencies of temporal constructions

As to the time relation, there are certain differences between translations from English and Russian. The temporal construction which is used when the action referred to is simultaneous with that of the main clause is clearly over-represented in translations from English, but the type used when the action has taken place earlier is used less in them than in original texts. In translations from Russian both types are used more than in texts originally produced in Finnish (Figures 5 and 6).
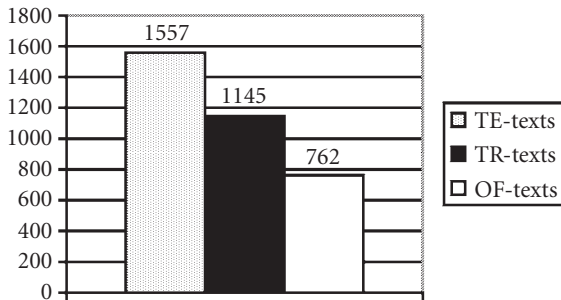

Figure 5. Frequencies of temporal constructions in the present tense
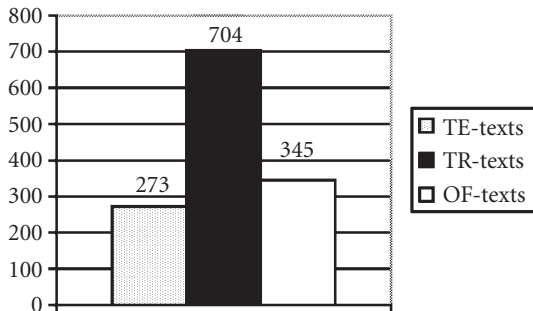

Figure 6. Frequencies of temporal constructions in the past tense

Different tendencies in texts translated from English and Russian may be influenced by actual frequencies of these theoretically possible constructions in these source languages. My own hunch is that past-tense gerunds are really common in the Russian language and more common than in English. The problem is the lack of evidence of real typicality of these verb forms in these two languages. As I already stated earlier, our intuition as to what is possible in a language is much more reliable than our intuition as to what is typical in it (see also Sinclair 1991: 39; Mauranen 2000: 138), and in this respect empirical results about the actual use of these structures in English and Russian are needed before any final statements can be made.

There is one interesting marked difference in the use of temporal constructions that concerns word order. In Finnish quite a free choice is available with respect to the position of the qualifiers of the temporal construction (although there are certain restrictions not specified here). For example in *autoa pestessään* (word for word translation: '*the car washing POSS.SUFF.*') the qualifier is in front position and in *pestessään autoa* (word for word translation: '*washing POSS. SUFF. the car*') it is in back position. As Table 4 indicates, in original Finnish texts qualifiers tend to be in front position more often than in translations. It is a well-known fact that in Russian gerund-structures and in English *ing*-participles, qualifiers are almost always (object without exception) in back position. This might influence the word order in translations.

Table 4. Position of qualifiers in temporal constructions with a possessive suffix

		Front position		Back position	
		f	%	f	%
TE-texts	N = 1112	285	25.6	827	74.4
TR-texts	N = 1297	298	23.0	999	77.0
OF-texts	N = 580	319	55.0	261	45.0

3.2.3   *The final construction*

The final construction is used to express the idea of aim or purpose. On the whole it is used in Finnish much less than referative and temporal constructions. Unlike the referative construction it has a clear straightforward equivalent in both Russian and English. The difference between these languages is that Finnish allows a choice between non-finite and finite forms and Russian and English most typically use non-finite forms (7).

(7) F a.   *Kiirehdin ehtiäkseni junaan.*   non-finite
        b.   *Kiirehdin, jotta ehtisin junaan.*   finite
     R   *Ja toropilas´ čtoby uspet´ na poezd.*   non-finite
     E   *I hurried (in order) to catch the train.* non-finite

As Figure 7 shows, translated texts are again strikingly different from the original Finnish texts but the tendency is the opposite of that in the referative construction: there are over twice as many final constructions in translations than in original Finnish texts.
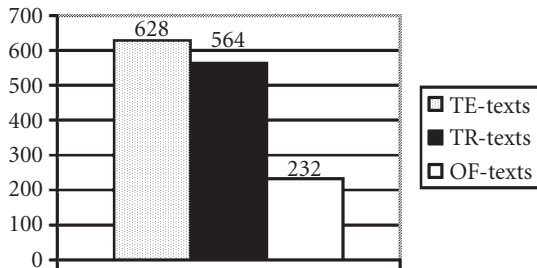

Figure 7. Frequencies of final constructions

It might well be that the infinite verb forms in English and Russian function as stimuli and result in over-representation of final constructions in Finnish translations. This is supported by Nousiainen (1982), who found that over 90% of final constructions in Finnish texts are translated as final (*in order*) *to-* constructions into English. She also argues that no other non-finite construction in Finnish has such a clear, straightforward equivalent in English. Quirk et al. (1972:753) state that in English, clauses of purpose are in the great majority of infinitivals.

In the use of the final construction there is evidence of a preference to place qualifiers in back position in translations more often than in original Finnish texts. The percentages are low (Table 5), but are still consonant with the results concerning the word order of temporal constructions.

Table 5. Position of qualifiers in final constructions

		Front position		Back position	
		f	%	f	%
TE-texts	N = 593	11	1.9	582	98.1
TR-texts	N = 532	13	2.4	519	97.6
OF-texts	N = 207	16	7.7	191	92.2

4. Discussion

The results show that translating does have an influence on the frequencies and distributions of Finnish non-finite verb forms. They suggest further that this influence has its source in the source language. Similar tendencies have been shown clearly also by two other Finnish researchers using corpus-based methods. Mauranen (2000) analysed word combinations – both collocations and multi-word strings – in Finnish translations and found that highly target-language-specific items tend to be under-represented in translations. Drawing on Reiss (1971), Tirkkonen-Condit (2000) studied some modal verbs that she calls unique, untranslatable items in Finnish (e.g. *jaksaa, malttaa, viitsiä, kehdata*), and found that they are used less in translations as compared to spontaneously produced texts. These are both lexical studies and now my results have shown that this kind of behaviour also holds for some syntactic structures. The linguistic choice between alternative, interchangeable expressions tends to produce different solutions in spontaneous writing and translating. This can be seen as evidence of the law of simplification: translators simplify by not using the resources of the target language according to its systemic possibilities as widely as the authors of original texts, but rather tend to keep close to the make-up of the source text and "forget" the alternatives available. In other words there are choices, but the variance in the way they are taken advantage of is smaller in translations than in original texts.

My main conclusion can thus be formulated as follows:

> Translations tend to under-represent target-language-specific, unique linguistic features and over-represent features that have straightforward translation equivalents which are frequently used in the source language (functioning as some kind of stimuli in the source text).

This means that the existence of a source-language stimulus raises the likelihood of using a corresponding construction in translation, and its absence reduces it. The hypothesis concerning the source-language stimulus is close to the idea of interference, which is of course not new. The notion of interference implies that translation reflects source-language features in a negative way. However, there are two basic differences between the "old" and the "new" way of looking at interference. First, statements about it have so far been made almost exclusively on the basis of the SL-TL relationship: what is new in the kind of research carried on by descriptive corpus-based translation studies is that evidence of interference can be seen on the grounds of target-language data only. Second, in the light of recent results it is important to see the impact

of the source language not as a negative phenomenon to be avoided but rather as a neutral, abstract and statistical, potentially universal phenomenon, just as the concept of *translationese* has recently become more of a neutral term referring to features that tend to distinguish translations from original texts. The hypothesis presented here is still just a hypothesis. It needs to be tested further on the basis of different kinds of corpora from different perspectives (for example analysis of contrastive differences between languages and real translation solutions using parallel corpora), and its universality should also be tested on the basis of large comparable corpora including different source and target languages.

References

Baker, Mona (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology. In honour of John Sinclair* (pp. 233–250). Amsterdam/Philadephia: John Benjamins.

Baker, Mona (1995). Corpora in Translation Studies: an Overview and Some Suggestions for Future Research. *Target, 7*, 223–243.

Baker, Mona (1996). Corpus-Based Translation Studies: the Challenges that Lie Ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in Language Engineering, In Honour of Juan C. Sager* (pp. 175–185). Amsterdam/Philadelphia: John Benjamins.

Baker, Mona (1999). The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics, 4*(2), 281–298.

Blum-Kulka, Shoshana (1986). Shifts of Cohesion and Coherence in Translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17–35). Tübingen: Gunter Narr.

Chesterman, Andrew (1997). *Memes of Translation: The Spread of Ideas in Translation Theory*. Amsterdam: John Benjamins.

Chesterman, Andrew (1998). Causes, Translations, Effects. *Target, 10*(2), 201–230.

The Corpus of Translated Finnish. Compiled at the Savonlinna School of Translation Studies, University of Joensuu, 1997–. Availability: Anna Mauranen, School of Modern Languages and Translation Studies in the Faculty of the Humanities of Tampere University.

Duff, Alan (1981). *The Third Language: Recurrent Problems of Translation into English.* Oxford: Pergamon Press.

Eskola, Sari (2002). *Syntetisoivat rakenteet käännössuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä* [Synthesising Structures in Translated Finnish. A Corpus-based Analysis of the Special Features of Finnish Literary Translations], [Publications of the University of Joensuu 30]. Joensuu: University of Joensuu.

Frawley, William (1984). Prolegomenon to a Theory of Translation. In W. Frawley (Ed.), *Translation: Literary, Linguistic, and Philosophical Perspectives* (pp. 159–175). London/Toronto: Associated University Presses.

Gellerstam, Martin (1996). Translation as a Source for Cross-Linguistic Studies. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast*: *Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4–5 March 1994* (pp. 53–62). Lund: Lund University Press.

Klaudy, Kinga (1995). Concretization and Generalization of Meaning in Translation. In M. Thelen & B. Lewandowska-Tomaszczyk (Eds.), *Translation and Meaning Part 3* (pp. 140–152). Maastricht: Hogeschool Maastricht.

Kohn, János (1996). What Can (Corpus) Linguistics Do for Translation? In K. Klaudy, J. Lambert, & A. Sohár (Eds.), *Translation Studies in Hungary*. Budapest: Scholastica.

Laviosa-Braithwaite, Sara (1996). *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation.* UMIST: Unpublished PhD thesis.

Mauranen, Anna (2000). Strange Strings in Translated Languages. A Study on Corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects* (pp. 119–142). Manchester: St. Jerome Publishing.

Nousiainen, Merja (1982). *Suomen kielen nominaalirakenteiden englanninkielisistä käännösvastineista* [On English translation equivalents of Finnish nominal structures]. Unpublished MA Thesis, University of Turku, Finland.

Øverås, Linn (1998). In Search of the Third Code: an Investigation of Norms in Literary Translation. *Meta, 43,* 571–588.

Pym, Anthony (1998). *Method in Translation History*. Manchester: St. Jerome Publishing.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, & Svartvik, Jan (1972). *Grammar of Contemporary English*. Harlow: Longman.

Rabin, Chaim (1958). The Linguistics of Translation. In A. H. Smith (Ed.), *Aspects of Translation* [Communication Research Centre, University College, London: Studies in Communication, Vol. 5] (pp. 123–145). London: Secker & Warburg.

Reiss, Katharina (1971). *Möglichkeiten und Grenzen der Übersetzungskritik: Kategorien und Kriterien für eine sachgerechte Beurteilung von Übersetzungen*. München: Hueber.

Schäffner, Christina (1999). *Translation and Norms*. Clevedon: Multilingual Matters.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Tirkkonen-Condit, Sonja (2000). In Search of Translation Universals: Non-equivalence or 'Unique' Items in a Corpus Test. Paper presented in the conference on Research Models in Translation Studies, 28–30.4.2000, Manchester, UK.

Toury, Gideon (1978). The Nature and Role of Norms in Literary Translation. In J. S. Holmes, J. Lambert, & R. von den Broeck (Eds.), *Literatures and Translation. New Perspectives in Literary Studies* (pp. 83–100). Leuven: Acco.

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

Toury, Gideon (1985). A Rationale for Descriptive Translation Studies. In T. Hermans (Ed.), *The Manipulation of Literature* (pp. 16–41). London: Croom Helm.

Toury, Gideon (1991). What are Descriptive Studies into Translation Likely to Yield apart from Isolated Descriptions? In K. M. van Leuven Zwart & T. Naaijkens (Eds.), *Translation Studies: The State of The Art. Proceedings of the First James S Holmes Symposium on Translation Studies* (pp. 179–192). Amsterdam: Rodopi.

Trosborg, Anna (2000). Translating Hybrid Political Texts. In A. Trosborg (Ed.), *Analysing Professional Genres* (pp. 145–158). Amsterdam: John Benjamins.

Vanderauwera, Ria (1985). *Dutch Novels Translated into English: The Transformation of a Minority Literature*. Amsterdam: Rodopi.

Weissbrod, Rachel (1992). Explicitation in Translations of Prose-Fiction from English to Hebrew as a Function of Norms. *Multilingua, II*(2), 153–171.

Zandvoort, Reinard Willem (1975). *A Handbook of English Grammar*. 7th edition. London: Longman.

Untypical patterns in translations

Issues on corpus methodology and synonymity

Jarmo Harri Jantunen

University of Joensuu

The aim of this chapter is to examine and further develop the corpus-based and statistical methods that have been used in investigations of universal tendencies in translations. It also attempts to test and further revise the hypothesis of untypical lexical patterning in translations (Mauranen 2000). The focus is on synonymous words and their lexico-grammatical patterning in three subcorpora of Finnish Comparable Corpus of Fiction (FCCF), which is a subset of the Corpus of Translated Finnish (CTF). Synonymous items have been studied due to their interesting and problematic nature in translations discussed already in both pre-corpus and corpus-based studies. The analyses are accomplished by applying a Three-Phase Comparative Analysis (TPCA), which is designed especially to analyse the source language influence. The TPCA and statistical procedures established that no clear and consistent evidence for a universal untypical lexico-grammatical patterning could be found, rather they provided support for a source language dependent tendency. Finally, it is suggested that generalizations concerning translation universals must be done carefully since investigations already carried out seem to show contradictory results, and since even the results in a one single study show partly different tendencies depending on either patterns or items that have been focused on.

1.   Introduction

Corpus-based analyses have assisted investigations into various research questions in translation studies. One of the areas where corpus studies have already had a great impact is the study of translation universals (Baker 1993).[1] So far, the hypothesised universal features studied by using methods of corpus linguistics are simplification (Laviosa-Braithwaite 1996, 1997; Laviosa 1998a; Jantunen 2001a) and explicitation (Olohan & Baker 2000). In her recent study,

Mauranen (2000) introduces a new candidate universal, namely untypical lexical patterning in translations: "it can – be suggested that lexical patterning which differs from that which is found in original target language texts might be a universal feature in the language of translations" (ibid. 136). Mauranen has reported that translations show both untypical frequencies of lexical items and untypical lexical patterning in translations. The former is manifested, for example, in the use of metatextual verbs (e.g. *haluta* 'to want to'), which are used more frequently in translations than in non-translations. The latter tendency is illustrated by connector *toisaalta* ('on the other hand'), whose lexical combinations in translations differ from those in non-translations. Furthermore, she notes (ibid. 128–129) that non-translated Finnish differs from translations from both English and non-English sources, which she interprets to indicate the independence of source language stimulus.

The aim of the present paper is twofold. First of all, it attempts to develop further a methodology that could be used to analyse universal tendencies and the influence of source language. The method utilised here will also be linked to the relevant methods used earlier in the context of corpus-based translation studies. Secondly, it aims to refine and test further the hypothesis of untypical patterning in translations. Since lexis and grammar are interrelated and indisputably dependent on each other (see e.g. Sinclair 1991, 1998; Hoey 1997), the present analysis concentrates not only on lexical strings, but also on grammatical patterning, and attempts in this way to complement the picture of the possible untypical patterns in translations. The hypothesis, based on the earlier findings, is stated as follows: *Compared to non-translated Finnish texts, translations into Finnish show (1) untypical frequencies of lexical items and (2) untypical lexical and grammatical combinations. The tendency takes place irrespective of the source language stimulus.*

This hypothesis is tested by studying the frequencies and lexico-grammatical association patterns of three synonymous Finnish degree modifiers, namely *hyvin*, *kovin* and *oikein* (all roughly meaning 'very'). Synonymous words are chosen, since in the field of translation studies, a consistent analysis of nearly synonymous words is – to my knowledge – still missing. The present paper is an attempt to bridge the gap between the investigation of synonyms and corpus-based translation studies, and to link the method and results of this study to the earlier findings on the use of synonyms in translated texts. The choice of synonymous words in general and the degree modifiers in particular is discussed in more detail in the next sections. Although a quantitative analysis forms the basis for this investigation, a qualitative approach is also included in order to give a comprehensive description for the question.

2.   Synonyms and the study of translations

2.1   Earlier studies on synonymity in translations

The study of synonymity and synonymous words have many aspects in common with studies of translations and translated language – that is the case especially in pre-corpus analyses of translation equivalents, but also recently in corpus-based studies of translation universals. Before machine-readable translational corpora were available, Blum-Kulka and Levenston (1983:119, 130–131) suggested that the use of common-level or familiar synonyms might account for lexical simplification in translations. In the 90's, at least two scholars shared their viewpoint: both Kohn (1996:48) and Laviosa-Braithwaite (1997:533) claim that the limited use of synonyms may be a sign of lexical simplification in translations.

The first corpus-based study that concentrates on both synonymity and simplification is Jantunen (2001a). In that study it was reported, contrary to earlier findings, that the range of synonymous words (amplifiers) is not narrower in translations; in some cases the range of synonymous degree modifiers is even wider in translations. Furthermore, it came out that translators do not tend to favour the most frequent synonym(s) at the expense of the other members of a group of synonyms. Mauranen (2000), however, has discovered that one of the synonymous expressions was overrepresented in translations, while another was favoured in non-translations. It seems that "a number of the differences between translations and originals [non-translations] involved different preferences in choosing between near-synonyms" (ibid. 138). These results are an interesting basis for further investigations because they seem to show quite opposite tendencies.

Synonyms can also be considered appropriate items in the analysis of untypical patterns in translation. It is claimed that each member of a group of synonymous words has distinct contexts in which they are used, and that this trait differentiates the word from its synonyms. Thus, we can analyse, firstly, what kind of contextual restrictions synonymous words have in language A, and secondly, whether the same restrictions and usage of synonyms are present in translations into the same language. In the next section, I shall describe some of the inherent characteristics of synonyms in more detail.

2.2 Lexical and grammatical patterning of synonyms

The meaning of synonymous words is similar with respect to their central semantic traits, but due to "minor or peripheral traits" (Cruse 1986: 267), synonyms are not interchangeable in all contexts. This is to say that synonyms are context dependent. According to Cruse (2000: 157), few, if any, synonymous words pass the test of absolute synonymity, meaning that lexical items would appear in exactly the same contexts. The contextual use of synonyms can be determined by linguistic and/or non-linguistic factors. The latter involves aspects such as register- (e.g. spoken language), dialect- (social or geographic) and style-specific (formal or colloquial) contextual restrictions. For example, the synonymous expressions *die* and *kick the bucket* have dissimilar ranges of use: *die* is more neutral and can be used in several contexts but *kick the bucket* is a more colloquial expression which could more presumably be found in slang or dialects than, let us say, in medical reports (for synonyms of *die*, see e.g. Cruse 1986). The linguistic factors in turn concern features which are not as obvious and visible as non-linguistic ones, that is, lexical and grammatical associations, which also determine and restrict the use of words. By lexical associations are meant the systematic co-occurrence patterns that a target word has with other words (see e.g. Biber et al. 1998: 6). This association is often called *collocation* and the adjacent words around target words *collocates* (see e.g. Firth 1968; Sinclair 1991). In other words, collocation refers to recurrent co-occurrences that a word has with its collocates within a given distance of each other, that is, in a pre-established span. The span can be determined by a structural unit (e.g. a sentence or entire text, see Kenny 2001: 90) but more commonly it is 'a short space' between a target word (a node) and its collocates, measured in words (Sinclair 1991: 170).

According to many scholars, only recurring or habitual co-occurrences can be considered as collocation. For example, Kjellmer (1987) counts only those associations that occur at least twice, whereas Kennedy (1991) puts the threshold at four occurrences – and in Jones and Sinclair's (1974) study the limit is set as high as ten occurrences. In addition to counting only the raw frequencies of collocations (as in Kenny 2001), the collocations are often analysed by using more or less statistical approaches. Mauranen (2000), for instance, has used relative frequencies (occurrences per million words) in comparison of lexical combinations in translations and non-translations and Biber et al. (1998) in analyses of synonyms. This norming of frequency counts is useful especially when corpora are not comparable in terms of length (Biber et al. ibid. 263). However, raw frequency counts or normed frequencies are

not able to tell much about the *strength* of co-occurrence between target word and its collocates. Another approach, namely tests for statistical significance that are based on observed and expected frequencies, can be used instead to measure the strength of association. Although statistical tests (see e.g. Stubbs 1995; Barnbrook 1996) or the whole statistical approach (Kenny 2001) include problems of their own, they are widely used in corpus linguistics to distinguish collocations that exist by chance from those whose co-occurrence is statistically significant. The statistical procedure used in this study is explained in more detail later in this chapter.

In its narrowest sense, collocation recognises only the lexical associations of nodes (Sinclair 1991: 170). However, the same kind of co-occurrences exist between node words and grammatical classes. These grammatical collocations are recognized as *colligations*. Colligations have, however, originally been defined as interrelations of grammatical categories, which thus concern categories such as word classes and sentence classes (Firth 1968: 181; see also Tognini Bonelli 1996: 74). In present-day corpus linguistics, however, colligation is understood to mean an association of a word, "seen as a unique lexical item rather than as a member of its class" (Tognini Bonelli ibid.), with grammatical categories (Hoey 1997: 8; Sinclair 1998: 15) or with a particular position in a sentence or text (Hoey ibid.; Kennedy 1991[2]). Both the contextual structures mentioned here (collocations and colligations) are crucial in the analysis of word meaning. As Carter puts it, meaning consists of several kinds of inter-relationships:

> – the meaning of a 'word' cannot really be adequately given without the fullest possible information concerning the place the word occupies and the contrasts it develops within a network of differential relations which includes patterns and ranges and the syntactic patterns which operate within particular ranges.
> (Carter 1987: 56)[3]

Corpus-based analysis of lexical or grammatical patterns suits particularly well the description of the use of nearly synonymous words (Biber et al. 1998). So far, however, corpus-based methods have not been widely used for this purpose. A few studies, though, are available. In their corpus-based presentation of language structure and use, Biber, Conrad and Reppen (ibid.) clarify the systematic differences in some groups of synonymous words. For example, nearly synonymous adjectives *big*, *large* and *great* have clearly different collocational association patterns in academic prose: *big* collocates most commonly with *enough*, *large* with *number* and *great* with *deal*. In another example, the synonymous verbs *start* and *begin* are studied, and a similar tendency is observed: *start* is more commonly used as an intransitive verb (*Blood loss started about the*

eighth day of infection . . .), while *begin* is used as a transitive verb (*Then I began to laugh a bit.*) In a study by Jantunen (2001b), Finnish adjectives *tärkeä* and *keskeinen* (both having a central semantic trait of 'important') show clearly different collocational and colligational association patterns. For instance, of the collocates that precede *tärkeä* ('important'), degree modifiers account for 11 per cent, but in the case of *keskeinen* ('important, central'), the proportion is only 3 per cent, to name but a few findings of the contextual associations.[4] The analyses listed here clearly show that the contextually dependent use of near-synonyms seems to differentiate them from each other.

3.   Methodology and data of the present study

3.1   Three-Phase Comparative Analysis (TPCA): a corpus-based method for investigating the impact of a source language in translations

The data for analysis consist of the Finnish Comparable Corpus of Fiction (FCCF), which is a subset of the Corpus of Translated Finnish (CTF) compiled at Savonlinna School of Translation Studies (for CTF, see Mauranen 2000; this volume). In translation studies, comparable corpus refers to a corpus which consists of subcorpora of both translated and non-translated texts (Baker 1995). The comparability usually means that texts are comparable in terms of genre, time of publication and possibly also in terms of text type and text length. The FCCF is composed of three subcorpora: (1) a corpus of non-translated Finnish (CNF), (2) a multi-source-language corpus of translated Finnish (MuCTF) and (3) a mono-source-language corpus of translated Finnish (MoCTF). The source languages in the MuCTF are: Indo-European languages like Dutch, English, French, German, Norwegian, Russian, Spanish, Swedish and Finno-Ugric languages like Estonian and Hungarian. For the source language in the MoCTF, I have selected English, which is an obvious choice for the reason that contemporary translations into Finnish are predominantly from English.[5]

The subcorpora of FCCF contain 0.8–1.0 million tokens each, which makes a total of 2.9 million tokens. Texts included in the data are full texts, and their total number is 50. They were published in 1995 or later, which means that they represent contemporary Finnish. As the name of the corpus indicates, the genre included in the corpus is fiction. Fiction is chosen because texts other than those of narrative fiction are rarely translated into Finnish from languages other than English. Fiction was then an obvious choice to
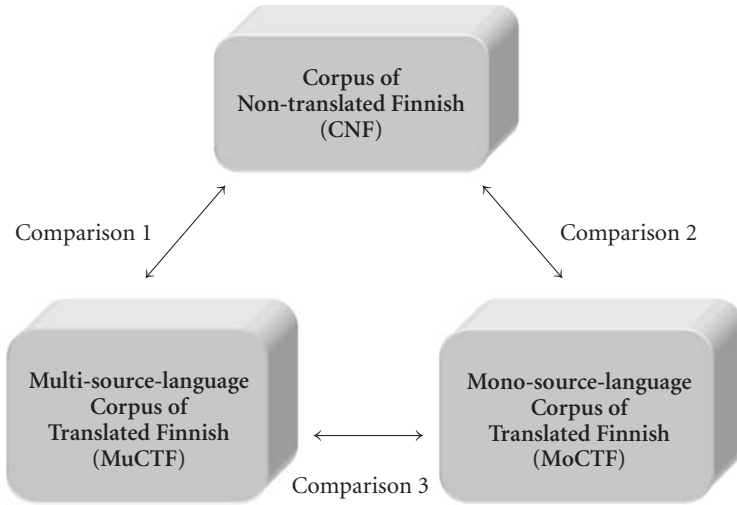
Figure 1. The Finnish Comparable Corpus of Fiction (FCCF) and The Three-Phase Comparative Analysis (TPCA)

make the compilation of the MuCTF possible. To avoid the idiosyncrasy of a particular text producer, only one text per writer or translator was included into the data. To describe the data briefly, the FCCF is a written, published, full-text and synchronic corpus that consists of both non-translational and translational subcorpora, the latter of which is divided into mono-source-language and multi-source-language subcorpora. (For a specific description of corpus typology, see Laviosa 1997.)

Why then two translational corpora? The aim of using two translational databases is not only to examine the possible impact of the source language on translated Finnish but also to study possible characteristics which translations from one particular source language could exhibit in comparison with translations in general. This will be tested through the Three-Phase Comparative Analysis (TPCA), which is illustrated in Figure 1.

Within this triple comparative perspective, phase one is formed of the comparison of CNF and MuCTF. The MuCTF is meant to represent so-called general translated Finnish, in other words, Finnish that has been translated from several source languages and which, presumably, does not reflect characteristics of any particular source language included in the corpus. In MuCTF, none of the source languages is dominant so it can be seen as a representative source of data the aim of which is to stand for translated Finnish in general. The

second phase, i.e. the comparison between CNF and MoCTF (comparison 2 in Figure 1) seeks to uncover the influence of one particular source language on translated Finnish. If the results of this phase are in line with the first comparison, we can presume that the source language does not influence on translated Finnish. On the other hand, if the findings are contradictory, the possibility of a source language impact cannot be excluded. The last phase (comparison 3), in turn, will complement the picture of translated Finnish by contrasting two translational corpora. In this phase, the MoCTF will be compared with the MuCTF to reveal whether the texts translated from one source language only may show dissimilar patterns from those retrieved from the MuCTF. That is to say, are translations from one source language different from translations in general in terms of lexico-grammatical patterning? If the outcome from both translational subcorpora turns out to be similar, the source language seems to have no impact on the patterning in translated Finnish, and vice versa.

The idea of investigating the impact of one particular source language is not unique, however. TPCA procedure can be said to be influenced by two earlier analyses, namely Laviosa's and Mauranen's. First of all, in her studies on simplification, Laviosa (Laviosa-Braithwaite 1996; Laviosa 1998a) also focuses on source language influence, although her overall methodology consists mainly of comparison of non-translational and multi-source-language corpora (see also Laviosa 1998b). In order to test the SL influence, she compares several translational subcorpora. Laviosa's model shows, however, significant differences compared to the method in the present chapter: while in TPCA, the aim is to examine the influence of *one specific* source language by using a *mono*-source-language corpus, Laviosa approaches the same question either by comparing *language-group-specific* SL corpora (e.g. Germanic with Romance languages) or *two-source-language* corpus (Italian together with Spanish) with two mono-source-language corpora (e.g. French) (Laviosa-Braithwaite 1996: 125, 129; Laviosa 1998a: 105–107). The latter type of comparison is nearly the same as the analysis performed in the present analysis, the former one, however, can be problematic, if an attempt is made to obtain information on the influence of one particular source language, although the lexical and grammatical make-up of related languages could reflect similar characteristics. However, Laviosa's primary aim has been to develop the methodology, and grouping of several source languages can be seen as a first stage towards an analytic research of SL impact (personal communication, 2001).

In line with Laviosa, Mauranen (2000) also aims to analyse the source language variable. However, the analyses clearly differ in terms of comparison procedure. Whereas Laviosa compares several (groups of) source languages,

Mauranen contrasts non-translated language primarily with translations from only one source language (English), and secondly with non-English sources (multi-source-language corpus). Although the latter comparison turns to be problematic due to the smallish quantity of translational material in the study (Mauranen ibid. 128, 135–136), the study attempts and succeeds in developing the methodology. In contrast to TPCA, Mauranen does not include English in the multi-source-language corpus; on the contrary, she uses *non*-English sources to study the source language variable. However, from a methodological point of view, English should be included among the SLs in a multi-source-language corpus, if its purpose is to represent translational Finnish in general instead of representing only translations other than those from English. To some extent, Mauranen also compares the two translational corpora in order to gain information on frequencies and combinations in several source languages. This aim and procedure are not, however, the most urgent questions in her study.

The last study that I would like to pick up, is Eskola's (2002) investigation of non-finite verb forms in two mono-source-language corpora (translations from English and Russian into Finnish) and in one corpus of non-translated language (original Finnish). By contrasting the frequencies and patterns retrieved from two mono-source-language corpora, Eskola's aim is to analyse the effect of one particular source language. Consequently, in the studies that have aimed to test and further revise the hypothesised universals of translation, several attempts have been made to obtain data on the impact of one specific source language. A summary of the different methods that have been exploited in analyses so far is presented in Table 1 below.

The software used in the present analysis is a concordance package Concord in WordSmith Tools (Scott 1998).[6] This program is used to generate concordance lines that include the node word (keyword) and its closest original

Table 1. Primary methods exploited for investigating the source language impact

Researcher	Procedure
Laviosa (1996, 1998a)	Comparison between a. language-group-specific SL corpora, and b. two-source-language corpus and mono-source-language corpora
Mauranen (2000)	Comparison of non-translations with a. English sources and b. non-English sources
Eskola (2002)	Comparison between two mono-source-language corpora

	2L	1L	node	1R	2R	
Silloin hän	sanoi	jotain	*hyvin*	kummallista.	#	Et sinä ole …
(Then s/he	said	something	very	strange.	#	You are not…)

Figure 2. The span. "L" strands for the left and "R" for the right side, the numbers mark the distance from the node. (An approximate translation in brackets.)

context. This Keyword in Context (KWIC) analysis is utilised to extract the immediate colligates and collocates of node words. After the relevant concordance lines (i.e. lines which include the node) are extracted from the corpus, they are sorted manually according to the word class in a given position. The span is limited to the space of two words to the left and two words to the right of the node. However, not only the words are counted but also the clause beginnings or ends, which could turn to be distinctive parameters in the analysis (see Figure 2).

3.2 Statistical procedures employed to analyse the similarity and difference

On the grounds that "one can never be entirely sure that the observed differences between two groups of data have not arisen by chance due to the inherent variability in the data" (Oakes 1998: 1), I have adopted a number of statistical procedures to avoid misconstructions of the data. The chi-square (x^2) test[7] (Butler 1985; Oakes 1998) is used to test the significance of observed frequencies in different subcorpora. Furthermore, statistical methods are used to test the significance and strength of collocations. To measure the significance, there are several tests available, of which z- and t-scores and Mutual Information (I) are the most commonly used (Barnbrook 1996: 94–100; for the range of tests, see also Oakes 1998).

According to Barnbrook, it can be difficult or even impossible to select one test that best evaluates the significance of the collocation in question (ibid. 101). This view is shared by Stubbs (1995), who claims that tests can be confusing and they must be interpreted with care. Both Stubbs and Barnbrook suggest that to achieve a balance between different tests it is probably better to use more than one statistical measure. In his analysis, Barnbrook (ibid. 100–101) reports that the three tests mentioned above provide different kinds of information on the significance of collocations: while both the z-score and the Mutual Information measures underline the significance of low frequency co-

occurring items, the t-score measure picks up collocations that are relatively frequent in the data. In the present analysis, I will partly follow Stubbs's (ibid. 40) suggestion of using both t-score[8] and I-measure[9] to test the significance of collocations. The significance is calculated by using a parallel ranking method, where each collocation is, firstly, sorted according to both scores (which gives two ranking lists: one sorted by I and the other by t-score). Secondly, the sorted lists are combined by summing the ordinals of each collocation into two lists. This method places the collocations in the final order of significance. To give an example, let us compare the significance of collocations *hyvin pian* ('very soon') and *hyvin pieni* ('very small') in non-translated Finnish. The ordinals that signify the rank in sorted lists are as follows (the real scores in brackets):

> *hyvin pian*: I (5.57) → 4. t (2.19) → 7.
> *hyvin pieni*: I (3.90) → 9. t (2.29) → 5.

It is clear that the tests emphasize the collocations differently: I picks up the collocation *hyvin pian*, whereas according to t-scores, *hyvin pieni* is the more significant collocation. To solve this problem, the ordinals are added up: 4 + 7 = 11 for *hyvin pian* and 9 + 5 = 14 for *hyvin pieni*. Thus, according to the two measures, of these two options the stronger collocation seems to be *hyvin pian*. Before this procedure, however, the collocations have already been filtered twice: first of all, only those collocates that are used by at least two writers or translators (i.e. collocates that exist in no less than two text files), and secondly, only collocations whose frequency is at least five ($\geq$ 5) are counted. This filtering is carried out in order to ignore idiosyncrasies and rare combinations or hapax legomena, which could be the result of creative use of language by a single text producer (see Kenny 2001). Finally, to test the significance of differences between the proportions of colligates, I have calculated the z-test for independent samples[10] (Butler 1985). In both tests, the significance is determined at the 5 per cent level ($p \leq 0.05$), which means that we can be 95 per cent sure that the results have not come up by chance.

4.   Quantitative analysis of the three most frequent boosters across corpora

The words chosen for a closer analysis are degree modifiers which premodify adjectives, adverbs, quantifiers and adposition structures (i.e. prepositional and postpositional phrases).[11] Degree modifiers are chosen for several reasons. First of all, the different groups of degree modifiers include a vast variety of

synonymous words. Moreover, Quirk et al. claim (1985: 441–453) that there are restrictions in the combinations of degree modifiers and grammatical classes. For instance, giving the example *The nail went right through the wall* they note that the number of intensifiers (here *right*) that can precede prepositional phrases (*through the wall*) is limited (ibid. 449). Degree modifiers can also be collocationally restricted, which is examined by Altenberg (1991) who states, for example, that many amplifiers tend to co-occur with words having a certain meaning (e.g. *utterly* co-occurs with words with negative sense). In line with Altenberg, also Bäcklund (1973), Paradis (1997) and Klein (1998) have found the same kind of colligational, collocational and semantic restrictions in the usage of degree modifiers in English and Dutch as well as in German. In Finnish, a comprehensive analysis of contextual restrictions of degree modifiers is lacking for the time being, although some efforts towards the description have been made (see Orpana 1988; Jantunen 2001c; Jantunen & Eskola 2002). In spite of the lack of thorough studies, however, we can expect that Finnish degree modifiers are correspondingly contextually restricted. Furthermore, degree modifiers are relatively frequent items and are used in texts regardless of the topic because they, at least the grammaticalized modifiers, are closer to function words than to content words (see Klein 1998: 27–28). Finally, degree modifiers do not typically vary in form, which, because FCCF is an unlemmatised corpus, makes the analysis straightforward.

Of the degree modifiers, boosters (i.e. modifiers that scale upwards from an assumed norm denoting a high but not extreme degree) are perhaps used most frequently, at least in English. This can be seen, for example, by comparing Tables 2.2–2.6 in Paradis (1997), and can partly be explained on the basis of exceptionally frequent use of the booster *very* (ibid. 34; see also Bäcklund 1973: 158).[12] According to A Frequency Dictionary of Finnish (Saukkonen et al. 1979), boosters are used frequently also in Finnish: of all degree modifiers booster *hyvin* ('very') is the commonest. Therefore, boosters – particularly the items *hyvin*, *kovin* and *oikein*, which are the commonest boosters in FCCF – are chosen for closer examination. The distribution of *hyvin*, *kovin* and *oikein* in FCCF is displayed in Table 2.

The frequency list shows that in every subcorpus of FCCF, the most frequent booster is *hyvin*, followed by *kovin* and *oikein*.[13] The rank frequency order is similar in every subcorpus, which indicates, firstly, that translated Finnish does not differ from non-translated Finnish in this respect, and secondly, that translations from English (MoCTF) do not differ from general translational language (MuCTF), either. It is, however, easy to see, that translations tend to differ from non-translations in another way. The total number of the de-

Table 2. Frequencies of *hyvin, kovin* and *oikein* in different subcorpora (raw counts and normed frequencies per 100 000 tokens)

	CNF		MuCTF		MoCTF	
	Raw	Normed	Raw	Normed	Raw	Normed
hyvin	352	36	709	66	555	70
kovin	176	18	414	39	306	38
oikein	121	12	157	15	156	20
Total	649	66	1280	120	1017	128

gree modifiers in CNF is considerably lower than those in both MuCTF and MoCTF: the total normed figures that indicate occurrences of all modifiers per 100 000 words are almost double in both translational corpora (120 and 128) compared to the normed figure (66) computed from CNF. A similar tendency can be found when we analyse the degree modifiers separately: in every case the normed frequency is bigger in translational corpora than in CNF – in cases of *hyvin* and *kovin*, the difference is especially clear.

To test and further define the hypothesis of untypical frequencies and patterns in translated texts, three chi-square tests (x^2) were performed. Tests were made in accordance with TPCA procedure: it was tested, firstly, whether the data for non-translations (CNF) differ from the data for MuCTF and secondly, whether translations from English differ from non-translations. Finally, MoCTF and MuCTF were compared. The calculated values of chi-square tests are as follows:

CNF vs. MuCTF: $x^2 = 16.11$
CNF vs. MoCTF: $x^2 = 3,81$
MuCTF vs. MoCTF: $x^2 = 4.92$

The critical value of x^2 at the 0.05 level of significance is 5.99. Since the value in the first test is greater (16.11) than the critical value, and the value in the second test smaller (3,81), we can conclude that there is a significant difference between the data for CNF and that for MuCTF but not between CNF and MoCTF. Thus, it seems that translations exhibit untypical frequencies of lexical items compared to non-translations but, more interestingly, the source language appears to influence the frequencies since the comparison between CNF and MoCTF is not in line with first comparison. Therefore, the hypothesis concerning untypical lexical frequencies in translations can be confirmed only partially. Finally, the third phase of TPCA shows that translations from one source language do not tend to exhibit untypical frequencies of lexical

items compared to translations in general since the value of the x^2 test (4.92) is under the critical value. Consequently, we can refine the earlier hypothesis and formulate a new hypothesis (concerning lexical frequencies) based on both statistical tests and Three-Phase Comparative Analysis: *Translated language tends to exhibit untypical frequencies of lexical items, but this tendency may be source-language dependent.* From the hypothesis it follows that untypical frequencies of lexical items are not considered to be a universal tendency in translations, rather a phenomenon that may well be influenced by a source language factor. This may seem surprising compared with the figures represented in Table 2 and particularly in the light of Mauranen's earlier findings. Both Table 2 and Mauranen (2000) rest on relative frequencies, which, in fact, show a very similar tendency, but which cannot be used alone to study and reliably test the similarities and dissimilarities of different language variants.

In the next sections, the focus will be on findings that concern lexical and grammatical associations of the degree modifiers. The presentation of the outcomes is divided into two main sections: firstly, the results that concern lexical combinations of all the three modifiers, and secondly, the results related to the grammatical combinations of one particular degree modifier, namely, *hyvin*.

5.   Lexical associations of synonymous modifiers *hyvin*, *kovin* and *oikein*

In the following section, the analysis of lexical associations will be limited to immediate right collocates, that is, the position 1R in concordance lines. These collocates function as syntactic headwords of the degree modifiers. Consequently, the collocates are likely to include adjectives (*hyvin* **väsynyt** 'very tired'), adverbs (*oikein* **hyvin**, 'very well'), quantifiers (*kovin* **paljon** 'very much') and prepositions (*hyvin* **lähellä** *kotiani* 'very near my home'). The distribution of the word classes of significant collocates is shown in Table 3.

First of all, the total number of significant collocates is clearly smaller in CNF than in the translational subcorpora. This must be partly due to the smaller number of modifiers in CNF, as displayed previously in Table 2. If the number of degree modifiers in a corpus increased, the number of different (significant) collocates would most likely also increase. Secondly, for every modifier, the proportion of each word class is broadly the same in every language variant. For example, the number of adjectives is almost equal to the number of adverbs. However, in one case there is a strikingly difference: the number and proportion of adjectival collocates of *hyvin* are

Table 3. The word classes that significant collocates of degree modifiers represent

	hyvin			*kovin*			*oikein*		
	CNF	MuCTF	MoCTF	CNF	MuCTF	MoCTF	CNF	MuCTF	MoCTF
adjectives	4	12	**17**	1	5	4	1	2	2
adverbs	3	11	**3**	1	4	5	1	1	2
quantifiers	2	2	1	2	3	2	–	1	1
adposition phrases	–	–	–	–	1	–	–	–	–
total	9	25	21	4	13	11	2	4	5

much larger in MoCTF than in CNF or MuCTF, and the proportion of adverbial collocates, in turn, is clearly smaller (in bold numbers in Table 3). It seems then that the source language can affect the lexical combinations, but the tendency for untypical lexical patterning is not consistent, because the untypicality is apparent only in the case of *hyvin*. Finally, the degree modifiers are dissimilar in terms of the number of significant collocates they get. This fact is less important in the context of untypical patterning in translations, but is especially important in the context of synonymy studies. It can be added, however, that the translation process does not seem to affect to the mutual ability of modifiers to obtain significant collocates.

The concrete collocates of *hyvin*, *kovin* and *oikein* in three subcorpora are shown in Tables 4, 5 and 6, respectively. The more comprehensive analysis of collocational combinations would require looking at the complete list of significant collocates; here I focus only on the 10 most significant collocates in each subcorpus. Each table gives the collocations and their approximate English equivalents. The collocates are sorted according to the two-significance-test-procedure described earlier (the most significant collocate is uppermost).

A striking difference can immediately be noticed between collocates in non-translations and translations in general: none of the collocates in the list of MuCTF occur in the list of significant collocates of CNF. The most significant collocation in CNF is *hyvin väsynyt*, whereas in MuCTF it is *hyvin tärkeä*. Other adjectival collocates in non-translations are *vanha*, *kaunis* and *pieni*; in MuCTF they are *erikoinen*, *yksinkertainen*, *vaikea*, and *vaarallinen*. In both lists, there are also adverbs, but they are different. In contrast to MuCTF, in CNF there are quantifiers, like *vähän* and *paljon*. In the second phase of comparison we can also find results that support the difference between translations and non-translations. The translations from English tend to have dissimilar collocations compared to non-translations; interestingly though, the

Table 4. Top 1R collocates of *hyvin* (raw frequencies in brackets)

CNF		MuCTF		MoCTF	
väsynyt 'tired'	(7)	*tärkeä* 'important'	(15)	*yksinkertainen* 'simple'	(8)
hitaasti 'slowly'	(8)	*harvoin* 'seldom'	(8)	*surullinen* 'sad'	(10)
hiljaa 'quietly'	(5)	*erikoinen* 'special'	(7)	*väsynyt* 'tired'	(7)
vähän 'little' (quant.)	(7)	*yksinkertainen* 'simple'	(7)	*kummallinen* 'strange'	(7)
pian 'soon'	(5)	*selvästi* 'clearly'	(8)	*kaunis* 'beautiful'	(10)
vanha 'old'	(6)	*vaikea* 'difficult'	(11)	*vaikea* 'difficult'	(8)
kaunis 'beautiful'	(5)	*varhain* 'early'(adverb)	(6)	*onnellinen* 'happy'	(7)
pieni 'small'	(6)	*lähellä* 'close' (adverb)	(7)	*hitaasti* 'slowly'	(7)
paljon 'much'	(5)	*varovasti* 'carefully'	(6)	*paha* 'bad'	(8)
		vaarallinen 'dangerous'	(7)	*pitkä* 'long, tall'	(10)

subcorpora share three collocations, namely *hyvin väsynyt*, *hyvin kaunis* and *hyvin hitaasti*. Since there are no quantifiers in the list, the range of word classes is narrower in the list head of MoCTF. These two phases of TPCA thus support the hypothesis of untypical lexical combinations in translations; the tendency seems to be unaffected by the impact of source language. However, we must keep in mind that we now discuss the overall tendency, not the actual word combinations, which as has already been seen may well be dissimilar in language variants.

In the final phase, in which we contrast translations from English to translations in general, we notice that in MoCTF *hyvin* is clearly being used to modify more adjectives than in MuCTF. As discussed above, the proportions of adjectives and adverbs were dissimilar in MoCTF and in translations in general (Table 3). When we focus on the list heads of significant collocates the same tendency can also be seen: among the 10 most significant collocates in MoCTF, there are nine adjectives and only one adverb (*hitaasti*) – and no quantifiers. Moreover, the list heads have only two collocates in common (*yksinkertainen* and *vaikea*); the other collocates in the top ten list are different from those retrieved from MuCTF. This analysis shows, then, that lexical patterns may distinguish translations from one particular source language from translations in general. However, we must make our conclusion keeping in mind that we have so far analysed only the 10 most significant collocates of *hyvin*. By extending the analysis beyond the list heads we could obtain a more complete picture of the collocational patterns of this specific degree modifier.

Table 5 below displays the top ten significant collocates of *kovin*. The analysis shows a somewhat different picture of the lexical bounds in language variants. Namely, three of the collocates in CNF also occur in MuCTF (*usein,*

Table 5. Top 1R collocates of *kovin* (raw frequencies in brackets)

CNF		MuCTF		MoCTF	
usein 'often'	(7)	*paljon* 'much'	(41)	*paljon* 'much'	(28)
moni 'many'	(11)	*kauan* 'for a long time '	(14)	*pitkälle* 'far '	(13)
paljon 'much'	(13)	*pitkään* 'for a long time'	(7)	*surullinen* 'sad'	(6)
tärkeä 'important'	(5)	*kauas* 'far' (adposition)	(5)	*hyvin* 'well' (adverb)	(12)
		vähän 'little'	(12)	*paha* 'bad'	(7)
		moni 'many'	(9)	*pitkään* 'for a long time'	(5)
		usein 'often'	(7)	*kauan* 'for a long time'	
		iloinen 'glad'	(5)	*hyvä* 'good'	(7)
		paha 'bad'	(6)	*pitkä* 'long, tall'	(13)
		hyvin 'well' (adverb)	(7)	*kauas* 'far ' (adverb)	(8)

Table 6. 1R collocates of *oikein* (raw frequencies in brackets)

CNF		MuCTF		MoCTF	
hyvin 'well' (adverb)	(26)	*hyvin* 'well' (adverb)	(36)	*hyvin* 'well' (adverb)	(31)
hyvä 'good'	(21)	*hyvä* 'good'	(28)	*hyvä* 'good'	(26)
		mukava 'nice'	(8)	*kovasti* 'hard' (adverb)	(8)
		paljon 'much'	(5)	*mukava* 'nice'	(7)
				paljon 'much'	(9)

moni and *paljon*) and one in MoCTF (*paljon*); and finally, six of the list head collocates in MuCTF also occur in MoCTF (*paljon, kauan, pitkään, kauas, paha* and *hyvin*). The lexical combinations seem to be then more similar across subcorpora than in the case of *hyvin*.

To obtain a broader picture of the collocational patterns of synonyms in different language variants, the significant collocates of *oikein* are listed in Table 6.

In CNF, there are only two significant collocates of *oikein*. The number of significant collocates is slightly bigger in translational corpora, as also in both cases discussed above. However, in contrast to *hyvin*, and partially to *kovin*, the collocates of *oikein* show a very stable patterning across the language variants. Both collocate pairs *oikein hyvin* and *oikein hyvä* occur in each subcorpus as the most significant combinations, and the mutual order of these two collocations is similar. Furthermore, the collocations in both translational data are also almost identical, except for the collocation *oikein kovasti*, which occurs only in translations from English.

This brief analysis of lexical associations shows that collocations may be very different in translations than in non-translations and furthermore, in

translations from one source language only compared with translations in general. An example of this is the case of *hyvin*. However, the situation is far more complicated, as we saw in the analyses of *kovin* and *oikein*. Contrary to *hyvin*, the degree modifiers *kovin* and especially *oikein* show less varied lexical patterning across the language variants. It looks that there is a continuum from less stable lexical collocations of *hyvin* to fixed collocations of *oikein*. This indicates that even very synonymous words may have different degrees of collocational variance. According to the Three-Phase Comparative Analysis of lexical associations, it is obvious that the collocational variance across language variants may be affected not only by (1) the language variant itself but also by (2) the actual words in a language, no matter how closely they are semantically or syntactically related to each other.

6.   Further analysis: grammatical associations of *hyvin*

Here I will focus only on one degree modifier, namely *hyvin*, which as we saw above exhibits the most varied lexical combinations across language variants. The aim of the following analysis is to complement the picture of variation by focusing on grammatical associations. In this case, I will focus not only on the 1R position but on the whole span from position 2L to position 2R (see Figure 2 above). The analysis is carried out by counting each of the word classes in a given position and then by calculating the proportion of each word classes.

I will start the analysis from the position 1R, which was already analysed above from the standpoint of collocations. The syntactic categories, that is to say, the colligates, which occur in position 1R are adjectives, adverbs and quantifiers as well as prepositional or postpositional phrases. The variance of word classes in this position is smaller than in other positions. This is due to the limited variety of the headwords that degree modifiers are able to premodify; in other positions there is more variety.

The first part of the TPCA of colligations (Figure 3 below) shows, to begin with, that *hyvin* clearly prefers adjectives in this position: their proportion is 66–75% of all colligates. The proportion of adverbs is obviously less (21–27%), and the proportions of quantifiers and adposition structures are very minor. Comparison of the distributions across subcorpora shows, first of all, that translations in general are very similar to non-translation: the proportions of each colligate are almost equal in this position. However, translations from English show a clearly different tendency: firstly, they differ from non-translations, which indicates an impact of the source language.[14] Hence, the
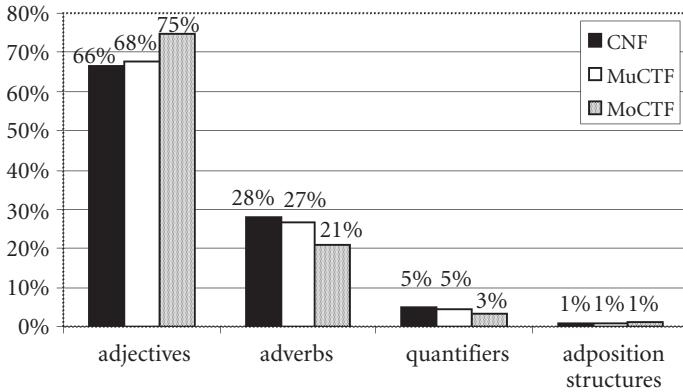
Figure 3. Proportions of 1R colligates of *hyvin* in subcorpora

two first phases of TPCA show that the proportion of head words of *hyvin* does not make a distinction between non-translations and translations, but this tendency seems to be dependent on source language impact. Secondly, the final phase of TPCA indicates that the translations from English show unequal colligational patterning compared to translations in general. This is the case for the proportion of adjectives, which make up a larger proportion in MoCTF (75% versus 68%), and in the case of adverbs, whose proportion is consistently smaller in MoCTF (21% versus 27%).[15] The proportions of quantifiers and adposition structures are, again, minor and also very similar. Thus, it seems that *hyvin* tends to colligate more strongly with adjectives and less strongly with adverbs in translations from English than in translations in general, which indicates that one particular source language may influence the grammatical structures of degree modifiers in Finnish.

Finally, I have completed the TPCA of the colligations of *hyvin* by analysing the whole span described earlier. In the analysis, the colligates were not only classified into word classes but the clause beginnings and ends were also counted. The comparisons of the language variants are displayed in Table 7, where only colligates whose proportion is significantly different are listed. At this point, the actual figures are not listed.

In the first row we see the results already analysed above; the other rows display the situation in positions two words to the right and both one and two words to the left. The first phase of TPCA shows very clearly that translations in general are very similar to non-translations. The proportions of colligates are equal (or not significantly different) in all but one position: in location 2R, the proportions of nouns and clause finals in MuCTF differ from those in CNF.

Table 7. The summary of distinctive colligates of *hyvin* according to TPCA

Position	CNF vs. MuCTF	CNF vs. MoCTF	MuCTF vs. MoCTF
1R	–	adjectives adverbs	adjectives adverbs
2R	nouns clause ends	–	nouns adverbs clause ends
1L	–	verbs	verbs quantifiers
2L	–	–	pronouns

The results of the second part of TPCA show a slightly different tendency: the proportions are now different in two positions, in 1R (as we already saw above) and in 1L. The number of colligates whose proportion is significantly different is, however, only a little larger than between CNF and MuCTF. The comparison does not then provide clear evidence for the source language impact, rather it implies that the SL does not clearly influence either *the number of colligates* whose proportion is different or *the number of positions* where the proportions are different. But more interestingly, there seems to be much a clearer difference between the two translational subcorpora. The last phase of TPCA reflects the specific nature of translations from English: analysis of the concordance lists shows that in every position of the span there occur at least one, usually two or more, colligates whose proportion is significantly different from those retrieved from MuCTF.

Remembering that we could find evidence for the influence of source language in terms of lexical combinations (at least in case of *hyvin*), the analysis of the colligates in the whole span of *hyvin* produces results that are in line with the earlier findings. Consequently, the analyses of grammatical and lexical patterning of *hyvin* appear to lend support to each other. However, we must remind ourselves that only the *proportions* of colligates are dissimilar across the language variants (when that was the case): the actual colligations, i.e. the grammatical combinations, turned out to be similar in every subcorpus. Thus, the colligation analysis showed only a quantitative, not qualitative, difference across language variants.

After summing up both the results of the analyses of collocations and colligations, we could formulate a new hypothesis concerning untypical lexical and grammatical patterning in translated language: *Translated language tends*

to exhibit untypical lexical combinations, but this tendency is dependent on the source language and the analysed words. Grammatical combinations tend to be similar in translations and in non-translations, but the impact of the source language on proportions of colligates cannot be excluded.

7.   Discussion

The present paper has aimed to analyse and complement the hypothesis introduced by Mauranen (2000) and, furthermore, to introduce and test a procedure that could be used to investigate the universal tendencies in translations. The method used here was Three-Phase Comparative Analysis (TPCA), which shares similarities with Laviosa's (Laviosa-Braithwaite 1996; Laviosa 1998a, b) and Mauranen's (ibid.) analyses, but which clearly differs in the way the impact of one particular source language was analysed. In TPCA, there are three comparative processes: Firstly, the comparison between non-translated texts and translations from several source languages, the aim of which was to find similarities and dissimilarities between non-translations and translated language in general. The second step was the comparison between non-translations and translations from one source language only, namely English. This phase aimed to test whether the results gained from the first phase could be interpreted as universal features or not. In the third and final phase, in turn, an attempt was made to clarify whether the texts translated from one source language exhibit characteristics different from those of translations in general. The analysis focused on three synonymous Finnish degree modifiers, that is the boosters *hyvin*, *kovin* and *oikein*, all meaning approximately 'very'. Synonymous words were chosen because it has been claimed in several studies that synonyms might be treated differently in source texts and their translations and on the other hand, also in non-translations of a given language and in translations into the same language.

Despite the fact that the primary aim of this chapter was to develop a methodology, the TPCA provided information on lexical and grammatical combinations both in non-translations and in translations and thus also offered information that could be used in research on wide-spread tendencies (universals) in translations. The results can be summarized as follows: no clear and consistent evidence for so-called translation universals could be found, but the results showed tendencies that might reflect the influence of the source language stimulus. To begin with, the overall frequencies seemed to show a clear SL independent tendency for overuse of degree modifiers in translations.

However, the statistical tests showed that a source language may affect overall frequencies, and the hypothesis of a universal tendency was rejected.

The analysis of collocations supported only partly the hypothesis of untypical lexical combinations in translations. First of all, translated texts, regardless of the source language, seemed to show dissimilar collocations compared to non-translated texts. This supports the hypothesis of untypicality. However, the *actual collocations* in texts from one source language turned out to be different from those in translations in general, which indicates that a source language may affect the lexical combinations. Perhaps surprisingly, this result was not consistent in the case of *all* synonymous degree modifiers, which indicates a clear influence of linguistic items on the results. The colligation analysis offered, again, results that do not support the hypothesis of a universal tendency. Although the translations exhibited almost the same grammatical patterning as non-translations, the translations from English differed in terms of the proportions of colligations.

Consequently, the analyses of lexical and grammatical associations as well as overall frequencies gave partly contrasting and rather more complex findings compared to the earlier investigations. Thus, it seems that the hypotheses need to be refined and studied more specifically. What I suggest is that quantitative hypotheses should be distinguished from qualitative ones. The present study suggests that quantitative and qualitative analyses give partially contradictory results. For example, although overall frequencies are partly untypical in translations (typicality of frequencies), combinations may be typical as well as untypical (typicality of patterning). More interestingly, it was the proportions (quantity) of items that distinguished language variants in the colligation analysis, not their actual range (quality).

The TPCA brings into the picture an important question about linguistic items which are focused to gain information on generalizations in translations. Although the present analysis and Mauranen's study use comparable methods and focus on lexical items, the results of these studies are not parallel. Moreover, as seen in this chapter, even the results of the analysis of words all belonging to one group of synonymous words may reveal contrasting results. Thus, the interpretations based on lexical combinations must be made very carefully before further and wider investigations have been carried out.

Apparently, it seems that the Three-Phase Comparative Analysis was a relatively useful and appropriate method for obtaining information about source language influence on frequencies and both lexical and grammatical patterning of degree modifiers in translations. However, some methodological points must be studied in the further analyses. For example, what is the impact

of one source language on the MuCTF – could one source language, or source languages belonging to the same language group, distort the distributions of collocational and colligational pattern, no matter how normally distributed the degree modifiers are in the corpus? And could we obtain different results for source language impact, if the source language was other than English in MoCTF? These questions cannot be answered in the current paper, but could be analysed in future research.

Notes

1. The concept of translation universal is used here knowing the criticism which it has met (e.g. Tymoczko 1998). It is used as a general concept referring to possible wide-spread tendencies in translations accepting that its manifestations might not concern all languages or language pairs at any given time and place.

2. See especially Tables 7.2 and 7.4.

3. According to Sinclair (1996:94), the investigation of word meaning requires not only the analysis of collocational and colligational patterns, but also the description of semantic preferences and prosodies, which also have a central role in language description. The contextual semantic categories are not, however, included in the present analysis.

4. Furthermore, also the morphological features (such as comparative and superlative forms) seem to differentiate these nearly synonymous words.

5. In 1999, the proportion of English among all source languages was as high as 69 per cent, and it seems that it will grow in the future (Minkkinen 2001).

6. WordSmith Tools is available at: http://www.lexically.net/wordsmith/

7. $x^2 = \Sigma (O - E)^2/E$, where O = observed frequency and E = expected frequency (Butler 1985:113; Oakes 1998:25).

8. $t = (O - E)/\sqrt{O}$, where O = observed frequency and E = expected frequency (Barnbrook 1996:97).

9. $I = \log_2(O/E)$, where O = observed frequency and E = expected frequency (Barnbrook 1996:98).

10. $z = (p_1 - p_2)/\sqrt{p_p(1 - p_p)(1/N_1 + 1/N_2)}$, where p = proportion of items and N = sample size (Butler 1985:94).

11. Degree modifiers have also been called *adverbs of degree* (Bäcklund 1973; Klein 1998), *intensifiers* (Bolinger 1972) or simply *adverbs* (Quirk et al. 1985). They can modify, at least in English, not only the word classes named here but also verbs, pronouns, and nouns (Quirk et al 1985; Altenberg1991).

12. Paradis (1997) has studied only spoken English. We must, of course, keep in mind that the distribution of degree modifiers may vary across genres and registers.

13. The boosters and their rank frequency order are partly different from the ones described in Jantunen (2001a: Table 6). This is due to difference in methodology and aims of the examination.

14. Differences are statistically significant: $Z = 2,699210$ ($p < 0.01$) for adjectives and $Z = 2,465150$ ($p < 0.05$) for adverbs.

15. Differences are statistically significant: $Z = 2,69259094$ ($p < 0.01$) for adjectives and $Z = 2,506196331$ ($p < 0.05$) for adverbs.

References

Altenberg, Bengt (1991). Amplifier collocations in Spoken English. In S. Johansson & A.-B. Stenström (Eds.), *English Computer Corpora. Selected Papers and Research Guide* (pp. 127–147). Berlin/New York: Mouton de Gruyter.

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

Baker, Mona (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target, 7*(2), 223–243.

Barnbrook, Geoff (1996). *Language and Computers. A Practical Introduction to the Computer Analysis of Language*. Edinburgh University Press.

Biber, Douglas, Conrad, Susan, & Reppen, Randi (1998). *Corpus linguistics: Investigating Language Structure and Use*. Cambridge University Press.

Blum-Kulka, Shoshana, & Levenston, Edward (1983). Universals of lexical simplification. In C. Faerch & G. Kasper (Eds.), *Strategies in Interlanguage Communication* (pp. 119–139). London/New York: Longman.

Bolinger, Dwight (1972). *Degree Words*. The Hague and Paris: Mouton.

Butler, Christopher (1985). *Statistics in Linguistics*. Oxford: Basil Blackwell.

Bäcklund, Ulf (1973). *The Collocation of Adverbs of Degree in English* [Acta Universitatis Upsaliensis]. Uppsala.

Carter, Ronald (1987). *Vocabulary. Applied Linguistic Perspectives*. London: Allen and Unwin.

Cruse, D. Alan (1986). *Lexical Semantics*. Cambridge University Press.

Cruse, D. Alan (2000). *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford University Press.

Eskola, Sari (2002). *Syntetisoivat rakenteet käännössuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä* [Synthesising Structures in Translated Finnish. A Corpus-based Analysis of the Special Features of Finnish Literary Translations], [Publications of the University of Joensuu 30]. Joensuu: University of Joensuu.

Firth, J. R. (1968). A synopsis of linguistic theory, 1930–1955. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth 1952–1959* (pp. 168–205). London: Longman.

Hoey, Michael (1997). From concordance to text structure: new uses for computer corpora. In P. Melia & B. Lewandowska-Tomaszczyk (Eds.), *Proceedings of the Practical Application of Linguistic Corpora Conference* (pp. 2–23). Poland: Lodz University Press.

Jantunen, Jarmo Harri (2001a). Synonymity and lexical simplification in translations: A corpus-based approach. *Across Languages and Cultures, 2*(1), 97–112.

Jantunen, Jarmo Harri (2001b). *Tärkeä seikka* ja *keskeinen kysymys*: Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? [*Tärkeä seikka* and *keskeinen kysymys*: What can corpus-linguistic analysis reveal about near synonyms?]. *Virittäjä, 105*(2), 170–192.

Jantunen, Jarmo Harri (2001c). Untypical lexical patterns: a universal feature in translations? Paper presented at the Conference Translation Universals – Do They Exist?, 19–20 October 2001, Savonlinna.

Jantunen, Jarmo Harri, & Eskola, Sari (2002). Käännössuomi kielivarianttina: syntaktisia ja leksikaalisia erityispiirteitä [Translated Finnish as a language variant: Untypical syntactical and lexical features]. *Virittäjä, 106*(4), 184–207.

Jones, Susan, & Sinclair, John (1974). English lexical collocations: A study in computational linguistics. *Cahiers de Lexicologie, 24,* 15–61.

Kennedy, Graeme (1991). *Between* and *through*: The company they keep and the functions they serve. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics. Studies on Honour of Jan Svartvik* (pp. 95–110). London/New York: Longman.

Kenny, Dorothy (2001). *Lexis and Creativity in Translation. A Corpus-based Study.* Manchester/Northampton: St. Jerome Publishing.

Kjellmer, Göran (1987). Aspects of English Collocations. In W. Meijs (Ed.), *Corpus Linguistics and Beyond. Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora* (pp. 133–140). Amsterdam: Rodopi.

Klein, Henny (1998). *Adverbs of Degree in Dutch and Related Languages.* Amsterdam/ Philadelphia: John Benjamins.

Kohn, János (1996). What Can (Corpus) Linguistics Do for Translation? In K. Klaudy, J. Lambert, & A. Sohár (Eds.), *Translation Studies in Hungary* (pp. 39–52). Budapest: Scholastica.

Laviosa, Sara (1997). How comparable can 'comparable corpora' be? *Target, 9*(2), 289–319.

Laviosa, Sara (1998a). The English Comparable Corpus. A Resource and a Methodology. In L. Bowker, M. Cronin, D. Kenny, & J. Pearson (Eds.), *Unity in Diversity? Current Trends in Translation Studies* (pp. 101–112). Manchester/Northampton: St. Jerome Publishing.

Laviosa, Sara (1998b). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta, 43*(4), 557–570.

Laviosa-Braithwaite, Sara (1996). *The English Comparable Corpus (ECC): A Resource and a Methodology for The Empirical Study of Translation.* PhD Thesis, UMIST, Manchester, UK.

Laviosa-Braithwaite, Sara (1997). Investigating simplification in an English Comparable Corpus of Newspaper Articles. In K. Klaudy & J. Kohn (Eds.), *Transferre Necesse Est. Proceedings of the 2nd International Conference on Current Trends in Studies of Translation and Interpreting* (pp. 531–540). Budapest: Scholastica.

Mauranen, Anna (2000). Strange strings in translated language: A study on corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies 1. Textual and Cognitive Aspects* (pp. 119–141). Manchester/Northampton: St. Jerome Publishing.

Mauranen, Anna (this volume). Corpora, Universals and Interference. In A. Mauranen & P. Kujamäki (Eds.) *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Minkkinen, Virpi (2001). *Book in the 1980s and 1990s. Publishing, distribution and reading of literature*. Helsinki: Statistics Finland.

Oakes, Michael P. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press.

Olohan, Maeve, & Baker, Mona (2000). Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures, 1*(2), 141–158.

Orpana, Terttu (1988). *Kuvaus vai kommentti. Tutkimus suomen kielen adjektiiviadverbien semanttisesta tulkinnasta* [Description or Comment? A Study on the Semantic Interpretation of Adjectival Adverbs in Finnish], [Opera Fennistica and Linguistica 2]. Tampere: University of Tampere.

Paradis, Carita (1997). *Degree Modifiers of Adjectives in Spoken British English* [Lund Studies in English 92]. Lund University Press.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, & Svartvik, Jan (1985). *A Comprehensive Grammar of the English Language*. London/New York: Longman.

Saukkonen, Pauli, Haipus, Marjatta, Niemikorpi, Antero, & Sulkala, Helena (1979). *A Frequency Dictionary of Finnish*. Porvoo: WSOY.

Scott, Mike (1998). *WordSmith Tools.* Oxford: Oxford University Press.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Sinclair, John (1996). The search for units of meaning. *Textus, 9*, 75–106.

Sinclair, John (1998). The lexical item. In E. Weigand (Ed.), *Contrastive Lexical Semantics* (pp. 1–24). Amsterdam/Philadelphia: John Benjamins.

Stubbs, Michael (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language, 2*(1), 23–55.

Tognini Bonelli, Elena (1996). *Corpus Theory and Practice*. TWC.

Tymoczko, Maria (1998). Computerized Corpora and the Future of Translation Studies. *Meta, 43*(4), 652–659.

Testing the basics

Translation-specific lexicogrammar?

Characteristic lexical and collocational patterning in Swedish texts translated from English

Per-Ola Nilsson
Göteborg University

This paper reports on an investigation of the Swedish grammatical word *av* ('of', 'by'), which is overrepresented in Swedish fiction translations from English in relation to Swedish non-translated fiction texts in the comparable part of *The English-Swedish Parallel Corpus (ESPC)*. The study also incorporated the most significantly overrepresented collocational patterns involving *av*. Through the investigation it became clear that the overrepresentation of *av* is general and significant, and that there is also significant overrepresentation of associated patterns involving lexical as well as grammatical words. The study further indicated that the patterns are mainly due to source language transfer.

1. Introduction

The purpose of this paper is to investigate translation-specific collocational patterning in Swedish fiction texts translated from English.[1] In the investigation, which is corpus-driven, the translation-specific distribution of the Swedish grammatical word *av* ('of', 'by') is described, along with the usage of constructions where the word is frequently found. English-Swedish cross-linguistic description is also made, in order to trace possible source items and constructions contributing to the specific distribution in the Swedish translated texts.

Collocation concerns the syntactic features of lexis in the sense that different lexical items have a smaller or greater likelihood of occurring together, as collocates (cf. Malmkjaer & Anderson 1991:301). A collocation has been defined as "a sequence of words that occurs more than once in identical form (...) and which is grammatically well-structured" (Kjellmer 1987, quoted in

Renouf & Sinclair 1991: 128). The definition of what has been termed "collocational framework" is slightly different; Renouf and Sinclair define a framework as "a discontinuous sequence of two words, positioned at one word remove from each other" (1991: 128). Thus whereas a collocation may be exemplified by the combination *a + feeling + of*, a framework may be exemplified by the discontinuous sequence *a + X + of*. Both types may include lexical as well as grammatical words, and in the case of frameworks it can be said that they have the potential for including different types of lexical words depending on the framework components. For instance, the framework exemplified potentially includes a noun because of the presence of the indefinite article.

Both lexical and grammatical words may have a specific distribution in translated texts. Distinctive distribution of lexical words has been pointed to as a possible universal of translation (Baker 1993: 245), and corpus studies of the lexical features of translated text do indicate that this can be a prominent feature of translated texts (Gellerstam 1989). Other studies indicate distinctive distributions of grammatical words in translated texts (Laviosa 1998). Proceeding from these observations, a logical next step is to investigate collocational patterns as a feature in translated texts. In the case of literary translation from English into Swedish, preliminary corpus study has indicated that such characteristic lexicogrammatical patterning may occur (Nilsson 2002). Further, as will become clear below, there is sometimes reason for discussing collocational patterning in slightly more abstract terms, in terms of *colligation*, which may be defined as co-occurrences involving individual words and a grammatical class of items.

2.   Material, aim and method

In a corpus-based investigation, the choice of the object of study is often one informed by intuition or previous research, or by a combination of the two. In a corpus-driven investigation, on the other hand, the linguistic material itself is allowed to decide what will be chosen for further study. Moreover, all of the material found in a specific investigation of this type is accounted for in a description, and in this sense the corpus-driven method is different from the corpus-based method not only through the choice of starting-point, but also in that corpus search results are not used selectively to illuminate a predefined theory (for a description of the corpus-driven approach, cf. Tognini-Bonelli 2002).

From the above it follows that although maintaining control of the status of the corpus material is important in a corpus-based investigation, it is of even more fundamental importance in a corpus-driven investigation. The potential of the corpus to yield results that are relevant to the research question is even more crucial in a corpus-driven study than in a corpus-based study for the reason that the researcher must take all of the results into account in description and analysis, and for the reason that theoretical statements are made on the basis of corpus evidence alone. For these reasons, it is of vital importance that the corpus is representative of the type of material of which it is proposed to be representative. In the case of an investigation of translated vs. original fiction texts, for instance, it is important that texts are representative of non-translated and translated text, respectively, and that "comparable" texts are as comparable as possible in terms of genre etc, although complete comparability can never be achieved, and although there is a multitude of problems associated with establishing comparability in translation corpora (cf. Laviosa 1997).

The corpus used for this study is the fiction part of the *English-Swedish Parallel Corpus* (ESPC), a combined comparable and aligned parallel corpus of English and Swedish original and translated fiction and non-fiction texts. The fiction corpus subcomponents used are each composed of 10–15,000 word extracts taken from 25 novels. The two comparable Swedish original and translated subcorpora contain 308,160 and 346,649 words, respectively. There are three text type categories in the fiction part of the corpus – general fiction, crime and mystery and children's fiction – and the English and Swedish texts are matched in terms of genre, although the match is not complete in some respects. The greatest difference between subcorpora is found in the children's fiction category, where there are five English originals and one Swedish original. In the other fiction categories, category differences are smaller (see Altenberg, Aijmer, & Svensson 2001).

An important issue pertaining to the textual status of the corpus is the fact that it is composed of extracts taken from the beginnings of novels, rather than of entire texts. It has been pointed out in other contexts that some textual features tend to be unevenly distributed in book-length texts (Sinclair 1991:19). This limits the range of studies that can be made using the ESPC, e.g. stylistic studies and studies of low frequency items. For other types of study, such as the present one, it can be assumed that there is an acceptable basic level of representativity, since this study is an investigation of high-frequency grammatical items that are likely to be fairly evenly distributed in longer texts. Further, the main focus in the investigation is on general features
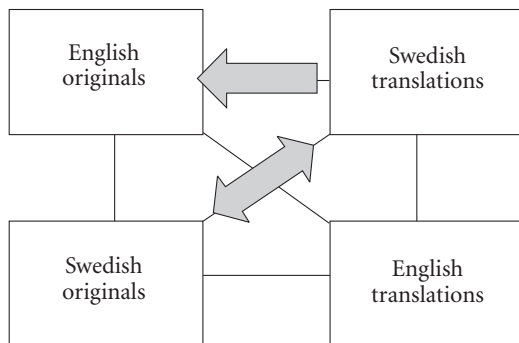
Figure 1. TL-TL frequency comparison and TL-SL qualitative analysis

in translation, and a corpus consisting of many extracts is better suited to capturing generalities than a corpus of the same size consisting of a smaller number of complete texts, where individual author and translator styles are likely to have greater impact on distributions.

The aim of this paper is to describe and briefly discuss the specific distribution of constructions involving the frequent Swedish grammatical word *av* ('of', 'by') in Swedish fiction texts translated from English. A range of collocational frameworks involving the word are described, one of them in some detail, and some attention is also devoted to specific cases of lexical words intervening in the frameworks.

The sense in which this study is corpus-driven is that frequencies are allowed to decide the object of study, on a general level as well as in more specific cases. The methodological starting-point of the investigation is to use differences in quantitative distributions between the Swedish comparable original and translated subcorpora in order to see what is quantitatively specific to the translated texts (diagonal arrow in Figure 1 above). The next step is to go back to the English originals of the Swedish translations to investigate the possible causes of specific TL distributions (horizontal arrow in Figure 1).

This means that the method is TL oriented in the sense that it involves starting from the TL rather than from the SL. The latter is the perspective more frequently opted for in earlier cross-linguistic studies of original texts and their translations. Much recent translation research, however, has a stronger focus on the translated text as an artefact of the target culture (cf. e.g. Toury 1995). The difference between the two perspectives is illustrated in Figure 2.

Method 1 results in a picture of a well-defined SL pattern being rendered as a paradigm of translational solutions in the TL. Method 2 gives a different
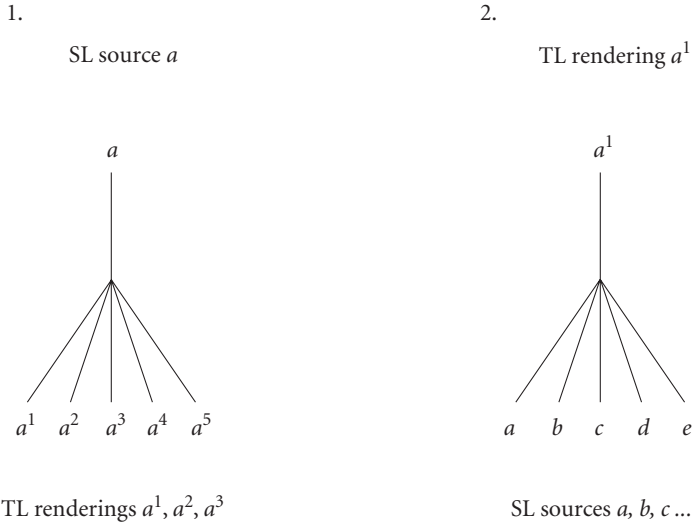
1.

SL source *a*

a

a^1 a^2 a^3 a^4 a^5

TL renderings a^1, a^2, a^3

2.

TL rendering a^1

a^1

a b c d e

SL sources *a, b, c* ...

Figure 2. SL and TL oriented cross-linguistic comparison

picture: When starting the analysis from the TL, the starting point is a well-defined TL construction – i.e., in the case of a translated text, a translational rendering – and a paradigm of SL patterns that give rise to it. Thus, where starting from the SL gives a picture of the multitude of possible translational solutions to specific types of source text problems, starting from the TL gives a picture of the multitude of types of source text problems that give rise to specific types of translational solutions. In other words, a TL oriented method is well suited to describing what in original texts contributes into giving translated texts certain specific features.

The procedure of investigation involved the following main steps:

1. A general quantitative comparison was made of original and translated TL texts so as to reveal any overall patterns of distinctive distribution in translated TL texts. The word *av* was the grammatical word showing the most significant total frequency difference, and was selected for further analysis.
2. The number of occurrences of *av* in each individual text file in the two TL subcorpora was recorded, and the generality of occurrence was then stated for each subcorpus in the form of a value expressing the standard deviation, i.e. to what extent there was variation around the medium value of occurrences of the word in the subcorpus, based on the frequency for each individual corpus text file.

3.   Using the same criterion of distinctive quantitative distribution as for *av*, as an individual grammatical word, the most significantly overrepresented TL collocational patterns of which *av* was a part were counted, and quantitatively compared with the corresponding patterns in the original TL subcorpus.

This initial quantitative TL verification was then followed by cross-linguistic qualitative analysis of SL-TL correspondences for selected TL collocational patterns and frameworks: The TL patterns with distinctive distribution emerging through steps 1–3 were used as a basis for further, cross-linguistic analysis, in order to reveal the actual sources of TL collocational patterns. (A further possible type of investigation, not carried out here, is to compare the TL collocational patterns in translated texts with the corresponding patterns in original TL texts; cf. Nilsson 2002.)

3.   Results

The Swedish grammatical word *av* is overrepresented in the translated Swedish comparable subcorpus, in relation to the subcorpus of original Swedish fiction. Table 1 shows this difference in distribution, as absolute frequencies and as percentages of the total number of words in each subcorpus.

The percentages for *av* in Table 1 reveal a highly uneven distribution in the two subcorpora – the frequency of the word in the translated texts approaches almost twice the frequency in the original texts, although the translated subcorpus as a whole is only around 12 % larger than the original subcorpus. The question arises how general this distribution is, i.e. to what extent the overrepresentation can be attributed to overrepresentation in individual texts.

A few basic statistical calculations reveal the following (based on the number of occurrences of the word *av* in each corpus text file expressed as a percentage of the total number of words in the individual file): The minimum and maximum percentages for individual files in the two subcorpora are 0.39 and 1.45 for the Swedish original texts and 0.75 and 2.21 for the Swedish translated texts. Expressed in terms of standard deviation, there is only a

Table 1. Distribution of *av* in Swedish original and translated fiction texts

	Sw. orig. fiction	%	Sw. trans. fiction	%
av ['of', 'by']	2,462	0.8	4,033	1.17

slightly higher degree of standard deviation for *av* in the translated texts: 0.33 against 0.24 in the original texts. These distributions may be illustrated in the form of a diagram – Diagram 1 shows the percentage levels for *av* in individual files in the Swedish original and translated subcorpora, arranged in ascending order of frequency:

Diagram 1. Distribution of *av* in individual corpus text files in the Swedish original and translated fiction parts of the ESPC
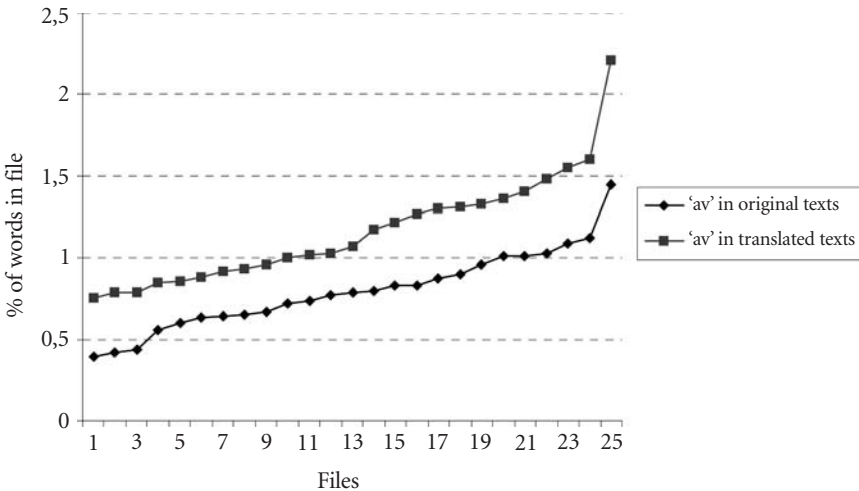
The diagram reveals that although the number of occurrences of *av* is generally higher in the translated files, the general distributions within the two respective subcorpora are in fact quite similar to one another. The word is more frequent in the translated texts, which is reflected as generally higher percentages for this subcorpus, but the word has an almost equally even distribution among translated texts as in original texts, which is reflected by the similarity of the two graphs in the diagram. On the basis of this, it can be said that the distribution of *av* is as general a phenomenon in the translated texts as in the original texts.

The next step after the establishment of the generality of the higher frequency of *av* in the translated texts is to describe collocations and frameworks incorporating the word in original and translated Swedish texts, and to define different subtypes. The definition of groups of such patterns can then lead to the definition of colligational patterns. For the collocations and frameworks described below, frequencies are much lower than for *av* as an individual word,

Table 2. Distribution of the colligational pattern LOCATIVE NOUN+*av* in Swedish original and translated fiction texts in the ESPC

Pattern		Sw. orig. fiction	Sw. trans. fiction
änden av	'the end of'	4	19
baksidan av	'the back of'	6	9
insidan av	'the inside of'	2	9
sidan av	'the side of'	14	27
mitten av	'the middle of'	2	14
utkanten av	'the edge of'	3	11
foten av	'the foot of'	1	7
hörnet av	'the corner of'	0	4
början av	'the beginning of'	8	18
slutet av	'the end of'	6	29
närheten av	'the vicinity of'	9	16
Total		**55**	**163**

and for this reason the distribution over individual files is not accounted for in this context.[2]

One example of a group of collocational patterns with a higher frequency in the translated texts is a specific type of nominal head: locative nouns, followed by *av* as a related structure word. Table 2 shows the distribution of these constructions in the two subcorpora.

The scope of analysis can be expanded further to the left to incorporate discontinuous triplets, frameworks where the noun + *av* patterns above may be one of several collocation types included here. Table 3 shows the distribution of a range of frameworks of the type *preposition + X + av*.

The combinations in Table 2 are in total around three times as common in the translated texts as in the original texts. The figures in Table 3 also reveal quite a significant degree of overrepresentation of frameworks of this type in the translated texts. Seeking an explanation for these differences of distribution, a relevant first question to ask is to what extent the translational renderings go back to structurally similar SL patterns and to what extent they are a result of several different types of SL structures converging, as it were, into one rendering (cf. Figure 2 above).

For the collocational category *änden av*, there is a high degree of structural correspondence between sources and translations – as could be expected for this type of phrase pattern, which is common and acceptable in Swedish: 15 of the 19 cases (cf. Table 2) can be said to exhibit structural correspondence. The range of source nouns lies semantically close to the noun in the translational

Table 3. Distribution of collocational frameworks of the type PREPOSITION + X + *av* in Swedish original and translated fiction texts in the ESPC

Pattern		Sw. orig. fiction	Sw. trans. fiction
i + x + av	'in' ...	106	165
vid + x + av	'by'/ 'at' ...	16	37
på + x + av	'on' / 'at' ...	75	122
med + x + av	'with' ...	44	82
mot + x + av	'towards' ...	9	14
från + x + av	'from' ...	13	21
till + x + av	'to' ...	21	36
under + x + av	'under' ...	5	22
efter + x + av	'after' ...	3	9
Total		**292**	**508**

rendering: *end* and *edge*, but also *side* and *bottom*. (The lexical words in the right contexts are words denoting spaces of various kinds – *room*, *garden*, etc, but also objects and some less tangible or more abstract notions, such as *line* or *journey*). Consider the following source text examples:

> at the **end of** the room
> to the **end of** the garden
> at the **far end of** the tunnel
> at the **other end of** the table
> at the **other end of** the train journey
> on the **other end of** the line
> to the **other edge of** the fence
> on the **far side of** the room
> from the **far side of** the lot
> on the **other side of** town
> at the **bottom of** the gardens

Thus, in the case of the combination *änden av* it is not a paradigm of SL patterns that causes TL overrepresentation, it is instead the sheer frequency of SL patterns that can be translated retaining the structure of the original. On the level of individual words, there is a paradigm of nouns – *end*, *edge*, *side* and *bottom* – that converges into the noun *änden* in translation. But from a structural or collocational point of view, there is more of a straight transfer, of the noun + *of* pattern, and it is the transfer of this colligational pattern that gives rise to collocational overrepresentation in the translated text.

In the case of the TL *vid + x + av* framework category, the distribution of corresponding source constructions is less uniform than in the case of the collocational pattern *änden + av*: 10 of the 37 instances (Table 3) can be classified as being renderings of non-corresponding structures.

The remaining instances, however, can be said to be cases of structural transfer. Consider the following selected source text examples:

> **at** the bottom **of** the off-ramp
> **at** the celebration **of** a marriage
> **at** the edge **of** a river
> **to** the edge **of** the cane field
> **on** the very edge **of** the sea
> **at** the end **of** the year
> **at** the end **of** the road
> **at** the foot **of** a tree
> **at** the foot **of** the Tor
> **at** the foot **of** my bed
> **At** the foot **of** the staircase
> **at** the group **of** farmers
> **upon** sale **of** the fourth stone
> **at** the side **of** the table
> **at** the very sight **of** a hypodermic needle
> **at** the thought **of** battery hens

As can be seen from the examples, some of the nouns from the earlier two-word collocation analysis turn up in the framework as well (e.g. *edge*, *end* and *side*). *At* is the most common grammatical word which appears in initial position in the source patterns, but there are also other grammatical initial position words (*to*, *on* and *upon*). As all these are translated as *vid*, it can be said that even when the structure as such is reproduced, there is a kind of convergence of grammatical initial position words that contributes to the overrepresentation of the target language pattern.

There is also a range of non-corresponding source structures that contributes to the overrepresentation of the framework in translations; cf. the following examples:

> rose and stood **next to** her father – *vid sidan av* ('by the side of')
> **Alongside** my real life – *vid sidan av* ('by the side of')
> **Apart** from Marie-Louise – *vid sidan av* ('by the side of')
>
> **while crossing** some Polish river – *vid övergången av* ('at the crossing of')
> felt actual physical nausea **at** such **sights** – *vid åsynen av* ('at the sight of')

> Hunt wheezed with cruel mirth **at** her black elastic belt – *vid åsynen av* ('at the sight of')

These source patterns may for instance be adverbials of different kinds, as in the first three examples. There is also a range of structures being translated as the TL framework with an intervening deverbal noun, as exemplified by the second group of phrases.

In summary of the cross-linguistic analysis of this framework, it can be said that the overrepresentation of the TL pattern is above all a result of a source language structure being transferred in similar form in translation (SL: PREPOSITION + NP + *of* – TL: PREPOSITION + NP + *av*). It is also to some extent a result of various other structures being rendered as the TL structure.

4.   Summary and conclusion

The main results of this study may be summarized as follows:

– The word *av* is significantly overrepresented in the translated subcorpus as a whole.
– The distribution of *av* is an equally general phenomenon in the translated texts as it is in the original texts.
– There is overrepresentation of collocational patterns as well as of collocational frameworks including *av*.
– The contrastive part of the investigation indicates that the cause of overrepresentation of the two TL patterns accounted for is a combination of the impact of the frequency of similar SL patterns and a range of other SL patterns, where source text frequency of similar SL patterns plays the largest role.

The generality of occurrence of the lexical item with which collocational patterns and frameworks are associated makes it reasonably safe to conclude that the patterns and frameworks themselves are also fairly generally distributed, although this awaits verification. As for the description of the treatment of individual collocational patterns and frameworks in translation, further qualitative study is necessary, one reason being that some SL patterns will have a translation equivalent close at hand in the form of a fairly fixed TL collocation, whereas for others there will be more of a paradigm of possible solutions.

Apart from supplying empirical results such as the above, the study also brings methodological issues to the fore, as well as questions regarding comparability. From a comparability point of view, the question is in what respects corpora can be said to be comparable if they are proposed to be comparable and are used as being so. In the case of the fiction texts used here, for instance, the differences for lexical collocations (e.g. *vid sidan av*; 'by the side of') may say more about culturally conditioned genre differences (in this case, perhaps, description of positions of objects in the world in certain genres of fiction) than about systemic linguistic contrast.[3] Collocational frameworks on the other hand, even if incorporating many possibly genre-related lexical patterns, may be slightly more interesting from the point of view of the linguistically oriented study of translations, since they reveal more about the ways in which basic and frequent lexicogrammatical source language patterning is treated in translation.

As for method, the exemplified way of using quantitative data for the definition of the specific linguistic object of study represents a connection between theory (hypothesis) and method in the sense that specific collocational patterning in translated texts is assumed to be a sufficiently typical and general feature of translated texts so as to be reflected on a global quantitative level even though it may not be a salient feature in any one translated text in isolation. This in turn leads on to the reception aspect of translation: Since patterns occur as generally in translated TL texts as in original TL texts as well as being more frequent in translated texts, they can reasonably be assumed to constitute a feature typical of Swedish fiction texts translated from English, at least within the time period and genre span covered by the corpus. On these grounds, the described patterns can be assumed to collectively contribute to the effect of a text being perceived as translated, along with other translation-specific patterning, collocational or other.

Notes

1. This study is being carried out as part of a project financed by *The Bank of Sweden Tercentenary Foundation*.

2. A calculation of individual distribution of items may however yield relevant information about the properties of specific translated corpus texts (cf. Nilsson 2002).

3. This "aboutness" of texts may in turn be contrasted with linguistic conventions of literary texts in a culture, such as for instance the usage of certain reporting verbs and formulae incorporating these (cf. Gellerstam 1996).

References

Altenberg, Bengt, Aijmer, Karin, & Svensson, Mikael (2001). *The English-Swedish Parallel Corpus (ESPC). Manual of enlarged version*. Online: http://www.englund.lu.se/research/corpus/corpus/espc.html.

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

Gellerstam, Martin (1989). Anglosaxiskt inflytande på svenskan. In *Orden speglar samhället*, 105–117. Stockholm: Allmänna förlaget.

Gellerstam, Martin (1996). Anföringens estetik. In O. Josephson (Ed.), *Stilstudier. Språkvetare skriver litterär stilistik* (pp. 12–29). Uppsala: Hallgren & Fallgren.

Kjellmer, Göran (1987). Aspects of English collocations. In W. Meijs (Ed.), *Corpus Linguistics and Beyond. Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora* (pp. 133–140). Amsterdam: Rodopi.

Laviosa, Sara (1997). How Comparable Can 'Comparable Corpora' Be? *Target, 9*(2), 289–319.

Laviosa, Sara (1998). Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta, 43*(4). Online: http://www.erudit.org/revue/meta/1998/v43/n4/003425ar.pdf.

Malmkjaer, Kirsten, & Anderson, James Maxwell (1991). *The Linguistics Encyclopedia*. London: Routledge.

Nilsson, Per-Ola (2002). Investigating characteristic lexical distributions and grammatical patterning in Swedish texts translated from English. In A. Wilson, T. McEnery, & P. Rayson (Eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World* (pp. 99–107). Munich: Lincom-Europa.

Renouf, Antoinette, & Sinclair, John (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: studies in honour of Jan Svartvik* (pp. 128–143). London: Longman.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Tognini-Bonelli, Elena (2002). Functionally Complete Units of Meaning across English and Italian: Towards a Corpus-driven Approach. In S. Granger & B. Altenberg (Eds.), *Lexis in Contrast* (pp. 73–95). Amsterdam/Philadelphia: John Benjamins.

Toury, Gideon (1995). *Descriptive translation studies – and beyond*. Amsterdam/Philadelphia: John Benjamins.

Explicitation

A universal of translated text?

Vilma Pápai
Széchenyi István University

This article reports on corpus-based investigation of explicitation generally referred to as one of the universal features of translation. It gives an account of the findings of a twofold analysis carried out on an English – Hungarian parallel corpus and a comparative corpus of translated and non-translated texts in Hungarian. The purpose is to reveal the regularities of both the translation process in terms of explicitation and the translation product in terms of text explicitness. The paper will argue that there is a close connection between explicitation and simplification, another candidate for translation universals.

1. Introduction

As all texts are shaped by the *particular aims* for which they were produced, the *particular context* in which they were composed, and by the *particular readership* to which they are addressed, translated texts must necessarily differ from non-translated texts. One of the main differences lies in the aim of text production. The ultimate goal of a writer is to produce a living, new text: "An author always wants to create sentences which have never existed in the given language before" (Esterházy 1996:182); a translator, however, renders texts created by someone else. In other words, the writer of a text seeks to achieve a formulation, a unique form of words, to fix and convey his matter, be it a story, relationship or idea. A translator, on the other hand, seeks to achieve a formulation to fix and convey the matter of another – a matter first conceived (and formulated!) in an idiom different from his own and that of his readers. As Baker puts it: "Translated text is normally constrained by a fully developed and articulated text in another language" (Baker 1996:177).

It is in the last decade that research into the nature of translated text, that is into its specific linguistic or discourse features, has gained new impetus mostly as a consequence of corpus methodology.

2.   Background

2.1 Explicitation

Explicitation is one of the features regarded as a universal of translated texts. Several studies have been carried out to test Blum-Kulka's hypothesis, which

> (...) postulates an observed cohesive explicitness from SL to TL texts regardless of the increase traceable to differences between the two linguistic and textual systems involved.   (Blum-Kulka 1986: 19)

In translation studies there have been two main approaches to challenge this hypothesis. Firstly, until recently research has been based on a comparison of a source text and a target text involved in translation. In consequence, findings have been articulated on the basis of contrastive analyses of a – what Toury calls – "series of (ad hoc) coupled pairs" (Toury 1995: 77), such as Dutch – English (Vanderauwera 1985), English – French and French – English (Blum-Kulka 1986; Séguinot 1988), Hebrew – English (Weissbrod 1992), English, French, Russian, German – Hungarian and vice versa (Klaudy 1993a, 1993b, 1996), English – Hebrew (Shlesinger 1995), and also Norwegian – English and English – Norwegian (Øverås 1996).

As a result, a number of textual features have been identified by drawing on theoretical and/or empirical research. Table 1 summarises the main characteristics considered to represent the special qualities translated texts display in comparison with non-translated texts as forms of a higher level of explicitness: longer texts, higher redundancy, stronger cohesive and logical ties, better readability, marked punctuation and improved topic and theme relation. In addition, this table also shows the views formed about the nature of explicitation as a strategy, the standpoints taken in the "a professional strategy vs. a by-product of language mediation" dilemma.

With the introduction of monolingual comparable corpora an entirely new approach to the investigation of translated text has emerged. This second approach can be called the "monolingual turn". Baker (1995: 234) formulates the merits of comparable corpora as follows:

> The most important contribution that comparable corpora can make to the discipline is to identify patterning which is specific to translated texts, irrespective of the source or target languages involved.

As scholars have adopted this alternative approach to the investigation of translated text (Laviosa-Braithwaite 1996; Kenny 1999; Olohan & Baker 2000), the text-to-text approach seems to be losing its importance (see also Laviosa 1998).

In their research, Olohan and Baker introduced the investigation of regularities in the use of optional elements in the language system. When investigating the Translational English Corpus (TEC) and the British National Corpus (BNC) they gave attention to the use of the reporting *that* in translated English texts.

2.2 Definitions and hypotheses

To discuss explicitation, we need to interpret this notion both in terms of the translation process and the translation product. For the purpose of the present research the following working definition of *explicitation* has been elaborated.

In terms of process, explicitation is a translation technique involving a shift from the source text (ST) concerning structure or content. It is a technique of resolving ambiguity, improving and increasing cohesiveness of the ST and also of adding linguistic and extra-linguistic information. The ultimate motivation is the translator's conscious or subconscious effort to meet the target readers' expectations. *In terms of product*, explicitation is a text feature contributing to a higher level of explicitness in comparison with non-translated texts. It can be manifested in linguistic features used at higher frequency than in non-translated texts or in added linguistic and extra-linguistic information.

With this in mind, I have formulated the following hypotheses: (1) in spite of the structural differences between the two languages the translation process from English into Hungarian involves explicitation strategies, (2) translated Hungarian texts show a higher level of explicitness than non-translated Hungarian texts, and (3) the degree of explicitness in scientific texts is higher than that of literary texts.

Table 1. Summary of nature & and forms of explicitation as represented in the literature

	Nature				Forms					
	Professional strategy	By-product of L. Mediation	Length	Redundancy	Cohesive ties	Logical ties	Readability	Punctuation	Theme – topic	Optional *that*
Blum-Kulka 1986		+	+	+	+					
Séguinot 1988	+*	+	+**		+			+	+	
Klaudy 1993		+	+		+		+			
Baker 1993, 1995, 1996		+	+		+	+				
Shlesinger 1995		+			+	+				
Toury 1995							+			
Øverås 1998			+		+					
Ishikawa 1998	+				+					
Olohan & Baker 2000		+			+					+

* a conscious strategy on the part of the editor

** some languages express the same message with a higher number of words than others (French compared to English)

3. Methods

3.1 Selection, structure and size of the corpus

The corpus assembled for this investigation (hereafter referred to as the ARRABONA corpus) consists of three sub-corpora put together from literary (L) and non-literary (N-L) texts (technical writing) written between 1969 and 1999:

1. the sub-corpus of original texts in English (OEC) is comprised of 8 texts written by British, American and Canadian writers,
2. the sub-corpus of their translations into Hungarian (THC) includes texts produced by professional translators and published by established publishing houses,
3. the sub-corpus of original texts in Hungarian (OHC) is made up of 8 comparable texts written in the same period (for the list of texts included see Appendix 1).

Figure 1 shows that these sub-corpora are designed to constitute a parallel (EHC) and a monolingual comparable corpus (HHC):
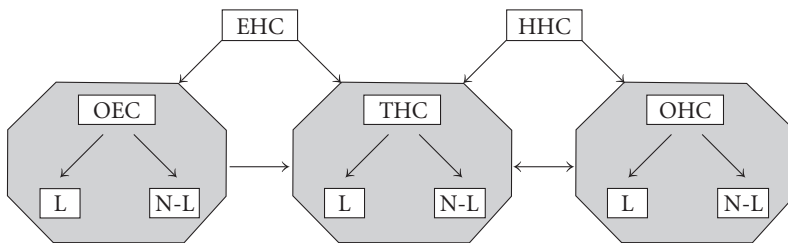
Figure 1. The structure of the ARRABONA corpus. A parallel corpus (EHC) & a comparable corpus (HHC)

The texts for this investigation were selected from a period spanning 30 years beginning in the late 70s and early 80s. The starting dates were mainly motivated by the intention to represent the period in which traditional Hungarian publishing standards were still at work. Existence of English translations of the Hungarian non-translated texts was another criterion for the selection of the original Hungarian works, which enables one to extend the analysis at a later time.

When selecting the texts for investigation, the overall intention and the main theoretical consideration was to achieve the highest possible variety

(Sinclair 1991:13–36): variety in terms of geography (British, American and Canadian authors), gender (male and female writers/translators) and status in the community. The lack of a translation-driven corpus imposed constraints especially in the selection of technical writing. Technical texts were included in the corpus in the belief that they contain a higher number of cohesive links than literary texts. Cohesive devices are the most frequently investigated text features (see Table 1), and are likely to provide insights into the nature of explicitation.

The three sub-corpora consist of the first 100 sentences of each text taken as representative of the texts as a whole in terms of the author's and the translator's style as well as being typical of the genre.

In total, the corpus contains *2,400* sentences yielding approximately *45,000* running words. WordSmith Tools (Scott 1998) was applied to align the sub-corpora of the EHC and also to carry out analysis on the HHC.

3.2 Methods

This corpus is designed to cater for two different methods of analysis. Through the investigation of the parallel corpus (EHC) I attempted to identify the explicitation strategies, i.e. types of shifts used by the translators when rendering English texts into Hungarian. In the procedure of identification the selection criterion was wider than that of Blum-Kulka's. Not only shifts in cohesion were included on the list but instances with additional linguistic and extra-linguistic information in addition to occurrences of ambiguous ST items rendered with disambiguated TT items. The guiding principle was to find instances of modification of ST, i.e. find steps towards an easy-to-understand, better structured, better organized and disambiguated text.

This manual analysis would also tell us about *the translation process* of the English → Hungarian translation direction. If we compare two languages, we will arrive at a conclusion that is restricted to the two languages in question and to the particular translation direction. If we intend to extend our claims, a monolingual comparable corpus as a tool will provide insight into the nature of translation and that of *the translation product.*

The second method of analysis serves this purpose. The procedure for this was entirely different: unlike in the first stage, where explicitation strategies were detected and analysed on a text-to-text basis, the second analysis looked at explicitness as manifested in textual features of 'large' bodies of texts. In addition, the investigation of the comparable corpus of translated and non-translated texts in Hungarian (HHC) was carried out with corpus methodology, mainly by using frequency data. The focus of investigation here was

to find out whether translated texts in Hungarian exhibited a higher level of explicitness than non-translated texts in Hungarian.

Technically, however, the two approaches were linked: five of the strategies identified in the first stage were taken in the second stage for further investigation and tested on the whole of the comparable corpus. Their selection was determined by the design of the corpus. As it is not annotated for parts of speech, only some of the strategies listed in Table 2, which lend themselves to frequency analysis, were selected for the second stage. Lexico-grammatical level, as a consequence, was entirely excluded. The strategies selected for further analysis: (1) addition and modification of punctuation marks (colon, semicolon, brackets), (2) addition of derivatives (*közötti* 'among' + [adjectival suffix], *belüli* 'in' + [adjectival suffix], *való* 'being'); (3) addition of conjunctions (*hogy* 'that', *aki* 'who', *ami* 'which, that', *amely* 'which, that', *pedig* 'but', *azonban* 'however, although, yet'); (4) addition of conjunctions and cataphoric reference (*az …, hogy, arr*…, hogy, ann*…, hogy* 'that' + demonstrative pronouns [with three different case endings]); and (5) addition of discourse particles (*csak* 'only, just', *még* 'still, yet', *is* 'as well, too', *például* 'for example', *így* 'so', *tehát* 'consequently'). These were treated as features of explicitness when comparing translated and non-translated texts of the HHC (Table 4). To sum up, the sequence of the methods used goes from detailed, close-up analysis to more general techniques of corpus methodology.

4. Results and discussion

4.1 The explicitation strategies

In the analysis of the English STs and their Hungarian translations 16 types of explicitation strategies were identified and categorized. Findings suggest that shifts occur on each level of language from the logical-visual level to the textual and extra-linguistic levels (Table 2).

The strategies identified at this stage of research were intended to cover most of the explicitation types, but they by no means present the full range of this translation strategy. They are sufficient, however, to provide a basis for the second analysis and also to show the variety of shifts on various language levels. Selected and limited examples of the strategies follow below; exemplifying each type of strategy would go beyond the scope of this study.

Table 2. Summary of explicitation strategies detected in the parallel corpus (EHC)

Levels	Shifts	Notes (reason/feature)
1. logical-visual relations	1. punctuation: addition and modification of punctuation marks 2. 1 S* → 2 Ss, 2 Ss → 1 S 3. explanatory conjunctions: e.g. *azaz (i.e.)*	conscious strategy &/or idiolect/language community style
2. lexico-grammatical	4. lexical repetition 5. grammatical parallel structures 6. filling elliptical structures 7. reconstructing substitutions 8. English pronoun → Hungarian noun	parallel structures
3. syntactic I.	9. derivatives I.: *lévő, való*** 10. derivatives II.: *közötti, belüli*	additions caused by structural non-equivalence in SL/TL
4. syntactic II.	11. addition of conjunctions 12. addition of cataphoric reference & conjunction	additions caused by differences in language economy in SL/TL (e.g. use of lower-grade devices of l. economy), conscious strategy: making explicit what was implicit in ST
5. textual & extra-linguistic level	13. lexical explanation 14. discourse-organizing items 15. situational addition 16. culture-specific items with added information	conscious strategy, language/genre conventions

*S – sentence
** - for explanation see 4.1.3.

4.1.1 *Shifts on the logical-visual level: punctuation marks*

Shifts in punctuation marks on the logical-visual level of text structure include instances of a) addition of a punctuation mark and b) replacing a punctuation mark with a stronger one. See example (1) for the former:

Table 3. Frequencies of punctuation marks in the EHC

Genre	Text	Colon	Semicolon	Brackets	Total
		E – H	E – H	E – H	E – H
L	CA	11 14	14 4	1 1	26 19
L	UP	2 7	11 14	1 1	14 22
L	PA	0 4	1 9	3 2	4 15
L	ME	1 9	2 7	0 0	3 16
Total		14 34	28 34	5 4	47 72
N-L	DA	10 25	10 7	4 7	24 39
N-L	EY	3 5	3 3	35 33	41 41
N-L	HA	2 11	0 5	10 10	12 26
N-L	HI	6 9	4 4	13 13	23 26
Total		21 50	17 19	62 63	100 132

(1) <DAenT, S 83> Paley here appreciates the difference between natural physical objects like stones, and designed and manufactured objects like watches.

<DAhuT, S 83> Paley itt azokat a különbségeket értékeli, amelyek a természetes fizikai objektumok (**mint a kő**) és a megtervezett és elkészített dolgok (**mint az óra**) között fennállnak.

Back translation: Paley here appreciates the difference that occurs between natural physical objects (**like stones**) and designed and manufactured objects (**like watches**).

Apart from handling extra remarks introduced by appositive *like*, the translator also makes the sentences more straightforward by inserting brackets. The summary of frequencies of punctuation marks in EHC (Table 3) reveals an overall rise in this feature, and moreover, reveals the differences in translators styles on the scale of preservation – alteration of the ST punctuation marks' patterns. Dashes were excluded from the analysis because of the essential difference in marking dialogues. The English language has a preference for quotation marks while Hungarian prefers dashes.

Shifts in punctuation marks into the stronger direction can be seen as "part of a subconscious strategy to make things easier, simpler, by making them more clear-cut" (Baker 1996: 182). It is also possible that the translators' ultimate aim initially is to make things clear-cut and more cohesive. Therefore, a simpler and easier-to-read text is the consequence of this strategy.

4.1.2   *Shifts on the lexico-grammatical level*

One of the cohesive ties is substitution. Shift in example (2) moves from one type of cohesive device to another, from substitution to lexical repetition. Although substitution "is a source of cohesion with what has gone before" (Halliday and Hasan 1976:90), the translator did not rely on its anaphoric reference and – probably for stylistic reasons – replaced it by a stronger cohesive tie.

(2)   <HAenT, S 56> As far as Kepler was concerned, elliptical orbits were merely an ad hoc hypothesis, and a rather **repugnant one** at that, because ellipses were clearly less perfect than circles.

<HAhuT, S 56> Kepler az ellipszispályákat alkalmi hipotézisnek tek-intette, méghozzá fölötte **visszataszító hipotézisnek**, mivel az ellipszis nyilvánvalóan tökéletlenebb a körnél.

Back translation: Kepler concerned elliptical orbits merely an ad hoc hy-pothesis, and a most **repugnant hypothesis** at that, because an ellipsis is clearly less perfect than a circle.

Reconstructing substitutions, i.e. replacing them by a noun head, however, does not appear to be a compulsory shift when translating from English into Hungarian, as shown in example (3):

(3)   <HAenT, S 47> Then two astronomers-the German, Johannes Kepler, and the Italian, Galileo Galilei- started publicly to support the Copernican theory, despite the fact that the orbits it predicted did not quite match **the ones observed.**

<HAhuT, S 47> Két csillagász: a német Johannes Kepler és az olasz Galileo Galilei nyilvánosan támogatni kezdte ezt a világképet, annak ellenére, hogy a Kopernikusz által megjósolt pályák nem minden esetben feleltek meg a **megfigyelteknek.**

Back translation: Two astronomers: the German, Johannes Kepler, and the Italian, Galileo Galilei started publicly to support this theory, even though the orbits predicted by Copernicus did not in each case match **the observed ones.**

The analysis of strategies on the lexico-grammatical level is based on Halliday and Hasan's typology of cohesive devices (1976) with the type of grammatical parallel structures established to cover several instances found in EHC.

Findings suggest that shifts occur in each type of cohesive devices in the English STs. They are replaced by different cohesive ties in the Hungarian

TT texts on the same level. Shifts including, for example, filling elliptical structures, reconstructing substitutions as well as lexicalising pronouns mostly result in lexical repetition, consequently lead to redundancy (Blum-Kulka). Why do translators often move from one type of cohesive tie to another?

If we take lexical repetition we will see a controversial device. On the one hand, translators tend to avoid lexical repetitions, in fact, this tendency is thought to be another candidate for translation universals (Baker ed. 1998: 288). On the other hand, as found in the data, translators end up using lexical repetitions in abundance to establish or strengthen cohesion in STs. While they want to create a clear and transparent target sentence, their aim can override the otherwise respected norm of translation, i.e. avoidance of repetition. This phenomenon, however, can be explained by the fact that "cohesion is part of the system of the language (. . .) and is built into the language itself" (Halliday & Hasan 1976: 5).

4.1.3   *Shifts on the syntactic level: derivatives*

As a consequence of the difference in structure of attributive complements: preference for postmodification in English, preference for premodification in Hungarian, translational Hungarian often uses words making left-branching of complement in noun phrases possible. Participles (very often empty *való* or *lévő* – *'being'*), semi-empty adjectives (*végzett* – '*conducted*') or postpositional adjectives *(közötti* 'among' + [adjectival suffix], *belüli* 'in + [adjectival suffix]) tend to fulfil this role. Example (4) involves a semi-empty adjective:

> (4)   <EYenT, S 8> In a fairly recent survey of academic psychologists in America, it was found that over three-quarters of them claimed to be cognitive psychologists!
> <EYhuT, S 8> Egy amerikai egyetemi pszichológusok között **végzett** újabb felmérésben azt találták, hogy több mint háromnegyed részük kognitív pszichológusnak tekintette magát.
>
> *Back translation:* In an among American psychologists **conducted** fairly recent survey it was found that more than three-quarters of them considered themselves cognitive psychologistst.

Hungarian postpositions (pl. *fák között* – *'trees among'*) with locative, temporal and other adverbial meaning can take suffixes and form postpositional adjectives like *között – közötti (közötti* – [között] + [i] – [postposition] + [adjectival suffix]). "Postpositional adjectives constitute the youngest Hungarian

part of speech and are representative of synthetic language structuring, whereas postpositives represent analytic language structuring" (Mátai 2002: 84).

(5)   <HIenT, S 2> The term Regional Economic Association (REA) defines collectively the various forms of economic integration among independent states.
<HIhuT, S 2> A regionális gazdasági társulás (REA) a független államok **közötti** gazdasági integráció különböző formáinak együttes definíciója.

Back translation: Regional Economic Association (REA) is a collective definition of various forms of 'independent states **among** +[**adjectival suffix**] economic integration'.

4.1.4   *Shifts on the syntactic level: conjunctions*

(6)   <HIenT, S 8> The dynamics of integration arise from increasing openness **and** political **and** economic interdependence among the participating countries.
<HIhuT, S 8> Az integráció dinamikája a részt vevő országok növekvő nyitottságának, **illetve** egymástól való kölcsönös gazdasági **és** politikai függésének eredménye.

Back translation: The dynamics of integration is a result of increasing openness **as well as** political **and** economic interdependence among the praticipating countries.

Example (6) involves a shift from one co-ordinator to another. Conjunct *illetve* ('as well as') changes the distribution of co-ordinated elements: instead of *openness and political and economic interdependence* the translator "works out" the relations: *openness 'as well as' political and economic interdependence*. There are probably two reasons for shifting the co-ordinator. First, the conjunct *and* lends itself to several interpretations:

> Semantically, linkage may be placed on a scale of cohesiveness ... *And* is the vaguest of connectives – it might be called a 'general purpose link', in that it merely says that two ideas have a positive connection, and leaves the reader to work out what it is.   (Leech & Short 1981 in Øverås 1998: 576)

Second, the x + (y + z) structure of the Hungarian phrase is probably easier to comprehend than the x +y +z structure in English.

4.1.5   *Shifts on the syntactic level: conjunctions + cataphoric reference*
One of the characteristic features of Hungarian relative clauses and *hogy* ('that')
clauses is cataphoric reference represented as an introductory pronoun in the
main clause. It is not necessarily part of the sentence but "its presence gives
the sentence a greater completeness" (Hell 1980: 157). Example (7) involves a
relative clause, and (8) constitutes a *hogy* clause:

(7)   <EYenT, S 80> If that is the case, then science may have practically no spe-
      cial features which elevate it above ancient myths or voodoo.
      <EYhuT, S 80> Ha ez így van, akkor a tudománynak gyakorlatilag nincs
      semmi **olyan** jellemzője, **mely** az ősi mítoszok vagy mágiák fölé emel-
      hetné.

      Back translation: If that is so, then science has practically no **such** special
      features **which** could elevate it above ancient myths or black magic.

(8)   <DAenT, S 19> The reader's reaction to this may be to ask, 'Yes, but are
      they really biological objects?
      <DAhuT, S 19> Az Olvasó reakciója valószínűleg **az** lesz, **hogy** megkérdezi:
      "Rendben, de vajon tényleg biológiai objektumok ezek?

      Back translation: The reader's reaction to this will probably be **that that**
      he/she asks, 'All right, but are these really biological objetcts?

4.1.6   *Shifts on the textual level: culture-specific items*
When the shared knowledge is different between two languages/cultures/
contexts the translator inserts a shorter or longer explanatory remark like
amerikai ('American'), that is the nationality of the publishing house, and *kiadó*
('publisher') in the second example.

(9)   <PAenF, S 7> A dozen years ago, a senior man from Knopf recognized
      his former prison guard inside the well-pressed suit of a Heibon-sha
      executive, stood staring at him for a moment or two, then threw his
      champagne into the startled Japanese face.
      <PAhuF, S 7> Tíz-tizenkét éve történt, hogy az **amerikai** Knopf egyik
      vezető munkatársa felismerte a Heibon-sha **Kiadó** valamelyik igazgatójá-
      ban azt a hajdani őrt – pedig jól-szabott öltönyt viselt –, aki annyit
      gyötörte valamikor a hadifogolytáborban; egy darabig csak állt és némán
      nézte, aztán pezsgőjét a megdöbbent japán arcába löttyintette.

      Back translation: It happened 10 or 12 years ago that a senior man from the
      American Knopf recognized in a Heibom-sha **Publishing House** executive

his former prison guard – althoug he was wearing a well-pressed suit –
who used to torture him so much in a POW camp, stood silently staring
at him for a moment or two, then threw his champagne into the startled
Japanese man's face.

The extensive/wide set of explicitation strategies identified in the parallel
corpus provides insight into the translation process in terms of shifts triggered
by a number of factors: the translators' conscious or unconscious strategy, or
the style of the translators or the language community, genre conventions or
translation norms, just to mention a few.

4.2 Shifts in explicitness

The comparison of translated and non-translated texts of HHC constitutes the
second analysis. Table 4 shows the distribution of frequencies of features of text
explicitness in the comparable corpus. The most relevant comparison is that
made between the original (O) and translated (T) Hungarian texts of HHC.
The data show that in 16 cases out of 20 (80%) the frequencies of features
investigated in translated text outnumber the frequencies in original text. The
most dominant difference was found in the case of derivatives *közötti* 'among' +
[adjectival suffix] and *belüli* 'in' + [adjectival suffix] with no instances in
original texts as opposed to 21 instances in translated texts.

Only in four cases do items conflict with the hypothesis: *való* 'being', *amely*
'which, that', *tehát* 'consequently', *is* 'as well, too' (Table 4). The most striking
results concern the use of *való*, an empty participle, and *tehát*, a conjunct of
consequence. There are 20 occurrences in the original texts for *való* as opposed
to 8 in translated texts, and 18 : 7 for *tehát*. The word *való*, the participle of
van 'to be' fulfils an important syntactic function: it makes the left-branching
of adjectives possible. As the Hungarian and English attributive complements
show a structural difference (see Appendix 2), we expect the frequencies of *való*
to be higher in texts rendered from an Indo-European language than in texts
produced by Hungarian writers. In other words, we do not expect writers to
use this item more often than translators do under constraints imposed by the
target text or the translation process itself or both; yet they do, with all but one
instance occurring in non-literary texts. As a result, this might be ascribed to
norms governing the use of these items for authors of technical writing.

The higher frequency of *való* in the non-translated texts, in fact, strongly
contradicts long-held professional views on this question. This unusual pat-
terning also applies to the conjunction of *amely* (a relative pronoun) and *is*

(discourse particle and also additive adjunct) but, of course, needs further investigation on a larger corpus.

We understand, therefore, that it is not, or not only, the translators who include elements thought to be translation-specific into technical texts. These patterns can be explained by the aim of technical writers who want to load a text with as much information as possible or, by their effort, conscious or not, to produce as clear a text as possible or, most probably by the influence of translated texts existing in the language community.

As for explicitness in the two genres we can observe higher frequencies in non-literary than in literary texts. However, only 65% of the cases (3 out 20) confirm the hypothesis, with the group of derivatives and conjunctions totally supporting Hypothesis 3 and with the items from the discourse particle group rejecting it. Explicitness of genres has to be investigated in more detail and corpora definitely can serve this aim.

To sum up, we can conclude that the frequency data in HHC provide evidence for the assumption that translated Hungarian texts show a higher level of explicitness than non-translated texts (Hypothesis 2). This can also mean that explicitation is likely to be a universal feature of translated texts, i.e. this set of data supports Blum-Kulka's hypothesis.

4.3 Type/token ratio in the comparable corpus

The type/token ratio is an indicator of lexical complexity as found on the surface of a text. The term *token* refers to the total number of running words, while the term *type* refers to the number of distinct word-forms in the text. The higher the percentage the more varied the vocabulary (Baker 1995; Munday 1998). The type/token ratio is considered to be very sensitive to the length of the text. I used the standardised type/token ratio because the texts of the ARRABONA corpus display the same number of sentences but reveal quite different word counts.

The findings of the statistical analysis indicate, as shown in Table 5, two tendencies. Firstly, that translations of the comparable corpus show a lower percentage type/token ratio than non-translations (58.15–63.29). This points to the conclusion that vocabulary used in the translated texts is less varied than that of the non-translated texts.

Secondly, non-literary texts of the comparable corpus show a lower type/token ratio than literary texts (57.69–63.74). This suggests that the vocabulary used in the non-literary texts is less varied than that of the literary texts.

Table 4. Frequency of features of text explicitness in the comparable corpus (HHC)

Level	Number	Group code	Feature	Literary			Non-literary			Original	Translated
				OHC	THC	Σ	OHC	THC	Σ	O	T
1.	1.	P*1	colon	21	34	55	56	50	106	77	84
	2.	P2	semicolon	1	34	35	10	19	29	11	53
	3.	P4	brackets	21	4	25	30	63	93	51	67
3.	4.	DR*1	közötti, belüli	0	4	4	0	17	17	0	21
	5.	DR2	való	1	0	1	19	8	27	20	8
	6.	C*1	hogy	80	121	201	86	176	262	166	297
	7.	C2	aki*	12	29	41	8	12	20	20	41
4.	8.	C3	ami*	30	38	68	28	39	67	58	77
	9.	C4	amely*	7	13	20	63	48	111	70	61
	10.	C5	pedig	6	8	14	13	18	31	19	26
	11.	C6	azonban	5	7	12	14	20	34	19	27
	12.	CR*1	az*…, hogy	5	26	31	28	54	82	33	80
4.	13.	CR2	arr*…, hogy	2	4	6	11	21	32	13	25
	14.	CR3	ann*…, hogy	2	2	4	2	9	11	4	11
	15.	DP*1	csak	18	37	55	18	30	48	36	103
	16.	DP2	még	24	30	54	10	23	33	34	53
5.	17.	DP3	is	59	59	118	51	33	84	110	92
	18.	DP4	például	1	1	2	10	19	29	11	20
	19.	DP5	így	11	23	34	18	16	34	29	39
	20.	DP6	tehát	5	0	5	13	7	20	18	7
			T/T	65.73	61.75	mean 63.74	60.84	54.54	mean 57.69	63.29	58.15

* P – punctuation, C – conjunctions, CR – cataphoric reference, DR – derivatives, DP – discourse particles
Relations marked with shade reject the hypotheses.

Table 5. Type/token ratios in the comparable corpus

	Sub-corpus	Original Hungarian	Translated Hungarian	*Mean*
1.	Literary	65.73	61.75	63.74
2.	Non-literary	60.84	54.54	57.69
	Mean	63.29	58.15	
	Difference	4.89	7.21	

At this point I wish to comment on the genres investigated. The type/token ratios of the non-translated texts indicate a 5% difference between the two genres, whereas this difference is 7% in the translated texts. The convergence of these parameters might suggest – as we hypothesed at the outset of the research – that translators of technical texts, in their effort to convey the information given by the ST as closely and clearly as possible, will inevitably use explicitation strategies more often than translators of creative literature. Explicitation strategies then lead to lexical repetitions, consequently to less varied vocabulary. In other words, this characteristic may well reflect the norm which governs genre expectations.

5.   Conclusions

The research reported in this paper purported to test the explicitation hypothesis and to examine whether translations have a higher level of explicitness than non-translations.

If we consider the structural differences between the two languages involved (the agglutinative Hungarian uses fewer words to express the same meaning than the analytical English, e.g. *I love you → Szeretlek*), translations from English into Hungarian would be expected to result in implicitation (making things more general, omitting linguistic or extralinguistic information of the ST) rather than in explicitation. With the 16 explicitation strategies, however, established in the parallel corpus, explicitation seems to be a strong tendency in the English – Hungarian translation direction.

The findings of the second analysis indicate a higher level of explicitness throughout the comparable corpus. Most of the frequencies support the hypothesis that the explicitness of the translations is higher than that of non-translated texts.

In addition, considering all the data, we can conclude that the third hypothesis should be rejected. The analysis did not provide evidence for the question concerning differences in genres. On the basis of the present findings we cannot claim that there is a clear-cut difference between literary and non-literary texts, as far as the linguistic items investigated are concerned.

As to the questions of type/token ratios, I am inclined to see the lower percentage in translated texts as a consequence of explicitation strategies. Apart from shifts on the logical-visual and textual – extra-linguistic levels (see Table 2), *all the shifts inevitably lead to lexical repetitions, consequently to simplification* in the vocabulary. For example, the addition of conjunctions, the addition of cataphoric reference + conjuncts, and the use of derivatives *add up, indirectly, to the number of repeated items*, i.e. the number of tokens, and therefore, *lessen the number of types*. Filling in ellipsis, reconstructing substitution, and replacing pronouns by nouns also contribute to a *lower variety of vocabulary*. This would allow the claim that the notion of explicitation seems closely linked to the notion of simplification in translation.

Summing up the analysis of the parallel and the comparable corpus we can conclude that, as the data suggest, in the period of 1969 and 1999 a translation norm was in play, according to which translators tended to adjust to target text standards and satisfy the target readers' expectations. On the whole, this is the ultimate function of explicitation strategies.

Appendix 1

Parallel corpus of English → Hungarian texts

Literary texts

Code	Author	Title	Translator	Title
UP	Updike, John 1977	*Marry me*	Göncz Árpád 1981	*Gyere hozzám feleségül*
CA	Capote, Truman 1980	*Music for Chameleons. Mojave*	Osztovics Levente 1982	*Mozart és a kaméleonok. Mojave*
PA	Porter, Anna	*The Bookfair Murders*	Bart István 1998	*Gyilkosság a könyvvásáron*
ME	McEwan, Ian 1999	*Amsterdam*	Tandori Dezső 1999	*Amszterdam*

Non-literary texts

Code	Author	Title	Translator	Title
DA	Dawkins, Richard 1986	*The blind watchmaker*	Szentesi István (1. fejezet) 1994	*A vak órásmester. Gondolatok a darwini evolúcióelméletről*
HA	Hawking, Stephen 1988	*A brief history of time: from the big bang to black holes*	Molnár István 1989	*Az idő rövid története*
HI	Hitiris, Theo	*European Community Economics*	Roboz András 1995	*Az Európai Unió gazdaságtana*
EY	Eysenck, Michael W., Keane, Mark T. 1990	*Cognitive Psychology. A Student's Handbook*	Bocz András 1997	*Kognitív pszichológia. Hallgatói kézikönyv*

Comparable Corpus in Hungarian

Literary texts

Code	Author – Translator	Title	Code	Author	Title
UP	Updike, J. 1977; Göncz Á. 1981	*Gyere hozzám feleségül*	KO	Konrád György 1969	*A látogató.* Budapest: Magvető Kiadó
CA	Capote, T. 1980; Osztovics L. 1982	*Mozart és a kaméleonok. Mojave*	MM	Mészöly Miklós 1984	*Megbocsátás.* Budapest: Szépirodalmi Kiadó

| PA | Porter, A. 1997; Bart I. 1998 | *Gyilkosság a könyvvásáron* | EP | Esterházy Péter 1990 | *Hrabal könyve.* Budapest: Magvető Kiadó |
| ME | McEwan, I. 1999; Tandori D. 1999 | *Amszterdam* | PO | Polcz Alaine 1991 | *Asszony a fronton.* Budapest: Szépirodalmi Kiadó |

Non-literary texts

Code	Author – Translator	Title	Code	Author	Title
DA	Dawkins, R. 1986; Szentesi I. 1994	A vak órásmester. Gondolatok a darwini evolúcióról	SI	Simonyi Károly 1978/1998	A fizika kultúrtörténete. Budapest: Akadémiai Kiadó
HA	Hawking, S. 1988; Molnár I. 1989	Az idő rövid története	BE	Bekker Zsuzsa 1978	Növekedési utak – dinamikus ágak. Budapest: Közgazdasági és J. K.
HI	Hitiris, Th. 1995; Roboz A. 1995	Az Európai Unió gazdaságtana	KJ	Kornai János 1983	Ellentmondások és dilemmák. Budapest: Magvető Kiadó
EY	Eysenck, M. W., Keane, M. T. 1990; Bocz András 1997	Kognitív pszichológia. Hallgatói kézikönyv	KF	Kozma Ferenc 1992	A menedzser közgazdasági szemlélete. Budapest: Közgazdasági és J. K.

Appendix 2

Premodification and postmodification
of Hungarian and English nouns

tágas	szoba		spacious room	
háromlábú	szék		three-legged chair	
esernyős	férfi		man	with an umbrella
nagyértékű	könyv		book	of great value
barnaöltönyös	férfi		man	in a/the brown suit
baloldalt lévő	ajtó		door	on the left
ablak mellett álló	lány		girl	(standing) by the window
művelésre alkalmas	föld		land	suitable for cultivation
egyezmény elérésére tett	erőfeszítések		efforts	(made) to reach an agreement
az az	ember,	aki éppen most érkezett	the man	who has just arrived

(Heltai – Pinczés 1993: 55)

References

Baker, Mona (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker, G. Francis, & E. Tognini-Bonell (Eds.), *Text and Technology: in Honour of John Sinclair* (pp. 233–250). Amsterdam/Philadelphia: John Benjamins.

Baker, Mona (1995). Corpora in Translation Studies: An Overview and Suggestions for Future Research. *Target, 7*(2), 223–243.

Baker, Mona (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175–186). Amsterdam/Philadelphia: John Benjamins.

Baker, Mona (1998). *Routledge Encyclopedia of Translation Studies.* London/New York: Routledge.

Blum-Kulka, Shoshana (1986). Shifts of Cohesion and Coherence in Translation. In *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*, J. House and S. Blum-Kulka (Eds.), 17–35. Tübingen: Gunter Narr Verlag.

Esterházy, Péter (1996). *Egy kék haris* [A blue-stock-ing]. Budapest: Magvető Könyvkiadó.

Halliday, Michael A. K., & Hasan, Ruqaiya (1976). *Cohesion in English.* London/New York: Longman.

Hell, György (1980). Object and subject *that* clauses in English and corresponding *hogy* clauses in Hungarian. In L. Desző & W. Nemser (Eds.), *Studies in English and Hungarian Contrastive Linguistics* (pp. 145–175). Budapest: Akadémiai Kiadó.

Heltai, Pál, & Pinczés, Éva (1993). *Fordítás és szövegértés a nyelvvizsgán* [Translation and reading comprehension in language examinations]. Budapest: Druck Kft.

Ishikawa, Luli (1998). Cognitive Explicitation in Simultaneous Interpreting. Cambridge: RCEAL University.

Kenny, Dorothy (1999). *Norms and Creativity: Lexis in Translated Text.* PhD Thesis. Manchester: UMIST. Centre for Translation Studies.

Klaudy, Kinga (1993a). Optional additions in translation. In *Translation the vital Link, Proceedings of the XIII. FIT World Congress, 2,* 373–381. London: ITI.

Klaudy, Kinga (1993b). On explicitation hypothesis. In K. Klaudy & J. Kohn (Eds.), *Transferre necesse est … Current issues of Translation Theory. In honour of Radó on his 80th birthday* (pp. 69–77). Szombathely: Dániel Berzsenyi College.

Klaudy, Kinga (1996). Back Translation as a Tool for Detecting Explicitation Strategies in Translation. In K. Klaudy, J. Lambert, & A. Sohár (Eds.), *Translation Studies in Hungary* (pp. 99–114). Budapest: Scholastica.

Laviosa-Braithwaite, Sara (1996). *The English Comparable Corpus (EEC): A Resource and Methodology for the Empirical Study of Translation.* PhD Thesis. Manchester: UMIST. Centre for Translation Studies.

Laviosa, Sara (1998). The corpus-based approach. *Meta, 43*(4), 473–659.

Leech, Geoffrey N., & Short, Michael H. (1981). *Style in Fiction. A Linguistic Approach to English Fictional prose*. London.

Mátai, Mária (2002). A névutók és a névutómelléknevek története [History of postpositions and postpositional adjectives]. *Magyar Nyelvőr, 126*(1), 72–87.

Munday, Jeremy (1998). A computer-assisted approach to the analysis of translation Shifts. *Meta, 43*(4), 542–556.

Olohan, Maeve, & Baker, Mona (2000). Reporting *that* in Translated English. Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures, 1*(2), 141–158.

Øverås, Linn (1998). In Search of the Third Code. An Investigation of Norms in Literary Translation. *Meta, 43*(4), 571–588.

Scott, Mike (1998). *WordSmith Tools.* Oxford: Oxford University Press.

Séguinot, Candice (1988). Pragmatics and the Explicitation Hypothesis. *TTR Traduction, Terminologie, Rédaction, 1*(2), 106–114.

Shlesinger, Miriam (1995). Shifts in Cohesion in Simultaneous Interpreting. *The Translator, 1*(2), 193–214.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Toury, Gideon (1995). *Descriptive Translation Studies and beyond*. Amsterdam: John Benjamins.

Vanderauwera, Ria (1985). *Dutch Novels Translated into English: The Transformation of a 'Minority' Literature*. Amsterdam: Rodopi.

Weissbrod, Rachel (1992). Explicitation in translations of prose-fiction from English to Hebrew as a function of norms. *Multilingua, 11*(2), 153–171.

Explicitation of clausal relations

A corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature

Tiina Puurtinen
University of Joensuu

The paper reports on a corpus-based study of clause connectives in translated and non-translated Finnish children's literature. The frequent use of clause connectives as explicit signals of clausal relations in translations might be one manifestation of the hypothesised translation universal referred to as explicitation. A one-million-word corpus of children's books both originally written in Finnish and translated from English into Finnish was used as research material to compare the relative frequencies of a number of connectives (conjunctions, specific adverbs, relative pronouns) signalling e.g. causal, temporal and postmodifying relations. The results reveal no clear overall tendency of either translated or originally Finnish literature using connectives more frequently, and thus fail to fully support the explicitation hypothesis. Nevertheless, in addition to some frequency differences, interesting differences were found between translations and originals in the functions and contexts of a few connectives.

1. Introduction

One of the hypothesised universals of translation is explicitation, which can refer either to making implicit source text (ST) information explicit in a translation, or to a higher degree of explicitness in translated texts than in non-translated texts in the same target language (TL). The studies by Vanderauwera (1985) and Blum-Kulka (1986), which address the first type of explicitation, show that target texts (TTs) tend to explicitate ST material e.g. by using repetitions and cohesion markers. More recently, the relation between translations and non-translations has started to attract more attention; Laviosa-Braithwaite

(1996) and Baker (1995, 1996) have presented hypotheses about the different frequencies of the optional *that*-connective in translated and original English texts, and Olohan and Baker (2000) report on a corpus-based study which shows that *that* is in fact more frequent in reported speech in translated than in original English.

This article focuses on particular explicit signals of clausal relations in children's literature translated into and originally written in Finnish, i.e. explicitation is here discussed as a potentially distinctive quality of translations in comparison with non-translated TL texts of the same type (as a "T-universal", see Chesterman in this volume). The question addressed below is whether clausal relations, or relations between propositions, are actually expressed more explicitly in translations, as the explicitation hypothesis suggests, by using a higher frequency of clause connectives such as conjunctions, specific adverbs and relative pronouns. An interesting, relevant study by Øverås (1998) has investigated a number of different cohesion markers in translations between English and Norwegian, and found that added connectives and replacement of connectives with more explicit ones are forms of cohesive explicitation in translations. Thus, in Øverås's study explicitation is examined as potential shifts between STs and TTs with no reference to comparable original TL texts, and therefore her findings are unfortunately not directly comparable to mine. Nevertheless, Øverås's research is interesting in that it includes similar cohesive ties as the ones in focus here, and the investigated texts represent fictional prose.

Mauranen's corpus-based study (2000) compares translated and non-translated Finnish texts, but the text type is different: academic prose and popular non-fiction. The analysis deals with text-reflexive (metatextual) expressions, including a number of connectors, and reveals that most connectors have roughly equal frequencies in translations and originals, with a slightly higher occurrence in translations. The main exception is *toisaalta* ('on the other hand'), which has over twice as many instances in Finnish originals as in translations; it has a tendency to combine with another connector (*mutta* 'but', *myös* 'also', *vaikka* 'although') in Finnish originals (cf. this result with my findings on *kun* in Section 4.1. below).

2. Explicitation of clausal relations

Since language use, including translation, is a matter of choosing between alternative ways of expressing meanings, and a particular choice is interesting

and meaningful in relation to the other alternatives that were not chosen, a brief look at the other options than using an explicit clause connective in a Finnish text is perhaps in order.[1] If, then, there is no explicit clause connective, there may be no other type of signal indicating the clausal relation either, but the relation must be inferred from the context. However, unlike e.g. the optional reporting *that* in English, a Finnish subordinative conjunction or relative pronoun cannot simply be deleted without making radical changes to the clause structure (see examples (1) and (2) below). Therefore, choosing this zero alternative in Finnish is likely to be a conscious strategy, whereas the English zero/*that* variation may be an unconscious one (see Olohan & Baker 2000: 143).

Instead of an explicit cohesion marker, a weaker signal, e.g. *ja* 'and', can be used. *Ja* is rather uninformative as it does not indicate the type of relation between clauses, unless the relation is simply additive. *Ja* can be employed as a kind of weak glue, to avoid creating a fragmentary, staccato rhythm by separating clauses with full stops. Finally, there are various other more or less implicit and rather complex realisations, referred to as nonfinite constructions (NCs). In the present context, the most interesting NCs are contracted clauses indicating temporal, referential and purpose relations, and premodified participial attributes equivalent to relative clauses. These constructions are very typical of Finnish texts and frequently used, although they sometimes tend to make the text "heavy", difficult to read and understand. They can be regarded as grammatical metaphors, which are marked, incongruent forms of encoding (Halliday 1994; Ravelli 1988). It is assumed that in English the typical, unmarked way of referring to an action, for example, is a verb, and using a noun instead is thus regarded as a marked, metaphorical expression. Similarly, qualities, which are usually realised by adjectives, can be expressed with nouns, and clausal relations, typically realised by connectives, can be expressed by nonfinite verb forms. What is considered a grammatical metaphor or a congruent realisation is a language-specific issue, but these basic ideas about English seem to be applicable to Finnish as well. (For the application of the concept of grammatical metaphor to Finnish texts, see Karvonen 1991 and Puurtinen 1993, 1995: 96–103.)

Example (1) shows two alternative ways of expressing a causal and a referential relation. The first, authentic version (Daniels 1998, trans. Jaana Kapari), includes two contracted clauses (in bold), whereas the second version (my formulation) signals the clausal relations with conjunctions.

(1)   *Hän katsoi aurinkoisesti äitiään **tietäen saaneensa jo anteeksi.***

<div align="right">(Daniels 1998: 7)</div>

literally: 'She looked sunnily at her mother knowing herself to have been forgiven already.'

*Hän katsoi aurinkoisesti äitiään, **sillä** hän tiesi, **että** oli jo saanut anteeksi.* (TP)

'. . . because she knew that she had already been forgiven.'

In example (2), the first alternative includes a premodified participial attribute construction and the second uses an equivalent relative clause beginning with the relative pronoun *joka* 'who', 'which'.

(2)   *Ja tiedäthän sinä, Mandy, että **lampaan kimppuun hyökännyt koira** on lupa ampua.*   (Daniels 1996: 59)

'And you do know, Mandy, that a-sheep-attacked-dog is allowed to be shot.'

*. . . että on lupa ampua koira, **joka on hyökännyt lampaan kimppuun**.* (TP)

'. . . that it is allowed to shoot a dog which has attacked a sheep.'

Relative clauses cannot, however, always be replaced with such compact structures, and therefore a writer or translator may not have several options to choose from.

NCs are likely to decrease the readability and speakability (ease of reading aloud) of a text, which are important qualities in a children's book. Cloze tests and reading aloud tests have shown that a high frequency of NCs makes a text significantly more difficult for children to both understand and read aloud fluently than the use of corresponding finite constructions with connectives (see Puurtinen 1995 for details). Surprisingly, previous research (Puurtinen 1995, 2003) has revealed that despite their negative effect on readability, NCs are relatively frequently used in translated children's literature. In English – Finnish translations of children's books published between 1940 and 1998, the frequency of NCs is significantly higher than in originally Finnish children's books from the same period. Moreover, NCs seem to have been becoming increasingly more common in translations since the 1970s. As NCs are associated with lack of connectives (as examples (1) and (2) show), it might be assumed that Finnish translations of children's books would have *lower* frequencies of clause connectives than non-translations. This assumption of course contradicts the explicitation hypothesis. The previous findings about the high frequency of NCs in translations can in themselves be interpreted as evidence contrary to the hypothesis. It is interesting that Eskola's corpus study (2002)

on NCs in adult fiction yielded partly different results: some NCs, i.e. temporal and purpose constructions, are overrepresented in Finnish translations of English and Russian fiction, whereas participial constructions are underrepresented. A plausible explanation for the different frequencies is the existence or non-existence of a formally equivalent structure in the ST, which may function as a trigger in the translation process. However, the intriguing difference in the use of NCs between translated adult and children's literature remains without explanation. Whether frequent use of NCs in children's fiction correlates with infrequent use of connectives is examined in the following.[2]

3.   Material and method

The material consists of translated and non-translated Finnish children's literature which forms part of the Corpus of Translated Finnish compiled at the Savonlinna School of Translation Studies (for compilation criteria, see Mauranen 2000: 122–123). The subcorpora used in this study are of approximately the same size (the corpus of translated children's literature has 593 000 words, the corpus of Finnish originals 500 000 words). The source language of all the translations is English, which is the dominant SL for both translated children's fiction and adult fiction in Finland. All texts were published between 1995 and 1998. The computer software used to retrieve the connectives is the WordSmith Tools program (Scott 1998).

The connectives selected for investigation are commonly used in all text types. The conjunction *ja* 'and' was excluded firstly because of its inexplicitness and secondly because of its function as a link not only between clauses but, even more often, between words and phrases, which would have meant a very time-consuming cleaning-up process to eliminate unlooked-for occurrences of *ja*. (The total number of *ja*, including the fused negative form *eikä* 'and not' was approx. 16 800 in each subcorpus.)

The investigated connectives are the following:

– the relative pronoun *joka* 'which', 'who' (9 cases, singular and plural forms)
– subordinative conjunctions

> temporal: *kun* 'when', *ennen kuin* 'before'
> purpose: *jotta, jottei(vät), että, ettei(vät)* 'in order to', 'in order not to'
> causal: *koska, kun* 'because'
> explicative: *että, ettei(vät)* 'that', 'that not'

>   conditional: *jos, jollei(vät)* 'if', 'if not'
>   concessive: *vaikka, vaikkei(vät)* 'although', 'although not'

–   coordinative conjunctions

>   adversative: *mutta, muttei(vät)* 'but', 'but not', *vaan* 'but'
>   explanative: *sillä* 'for'

–   adverbs

>   causal: *siksi* 'therefore'
>   adversative: *kuitenkin* 'however'

In addition to the basic forms, the fused forms composed of the conjunction and the negating word *ei*, as well as the plural forms with the suffix *-vat/vät* were searched for. The following discussion will be restricted to those findings that seem somehow interesting, not necessarily in terms of frequency differences as such, but functions and contexts of use.

4.   Results

Table 1 shows the frequencies of the connectives per 100 000 words. (Of the 18 different forms of the relative pronoun *joka*, only the ones with clearly different frequencies in Finnish originals and translations are presented.)

There seems to be no clear overall tendency of either subcorpus favouring connectives more than the other. Instead, some connectives are more frequent in Finnish originals (*jo(i)ssa, vaikka, vaan, kuitenkin*), others in translations (*jo(t)ka, kun, jos, ennen kuin, jotta*), and a few connectives have roughly equal frequencies in both subcorpora.

An attempt was then made to find potential explanations for the frequency differences, such as using a particular connective in partly different functions in translations and originals. An examination of the contexts surrounding such connectives produced some interesting observations, but in a number of puzzling cases even a closer look failed to reveal possible reasons for the discovered differences. For instance, the higher frequencies of *ennen kuin* and the nominative singular and plural forms of *joka* in translations could not be attributed to contextual or functional aspects. Only those few cases where the context turned out to be more helpful are discussed in more detail below.

Table 1. Occurrences of connectives per 100 000 words in Finnish originals and Finnish translations

	Originals	Translations
jo(t)ka 'which', 'who'	319.8	357.2
jo(i)ssa 'in which', 'where'	75.0	45.5
että, ettei(vät) 'that', 'that not',		
'in order to', 'in order not to'	1129.5	1155.7
kun 'when', 'because'	661.6	809.3
jos, jollei(vät) 'if', 'if not'	252.0	268.7
vaikka, vaikkei(vät)		
'although', 'although not'	126.3	111.3
koska 'because'	69.6	76.2
ennen kuin 'before'	46.1	95.3
jotta, jottei(vät)		
'in order to', 'in order not to'	11.2	56.7
mutta, muttei(vät) 'but', 'but not'	786.1	796.2
sillä 'for'	84.0	81.5
vaan 'but'	41.1	33.4
kuitenkin 'however'	58.1	46.7
siksi 'therefore'	22.3	21.2

4.1 Connectives more frequent in translations

The temporal conjunction *kun* 'when', which is sometimes also used causally, is considerably more frequent in translations (809.3 vs. 661.6 occurrences per 100 000 words). It seems to occur in word combinations such as the ones shown in Table 2 (the figures for individual items in Table 2 indicate absolute, not relative frequencies).

These time expressions are clearly more common in translations, although there is no apparent reason for avoiding them in Finnish originals, as all of them are perfectly acceptable Finnish expressions. Nevertheless, at least most of them are likely to have been triggered by a formally more or less equivalent English phrase (*juuri kun* ← just when, *nyt kun* ← now that, *sillä/samalla hetkellä kun* ← at the same time as), which suggests that, when possible, translators tend to translate the ST expression literally into Finnish.

The explicative and purpose conjunction *että* 'that', 'in order to' is more frequent in the translation subcorpus. However, its positive/neutral and negative forms (*että* can also appear in a negative clause separate from the negating word *ei*, e.g. *että hän ei ollut* 'that he was not'; *ettei(vät)* merges the conjunc-

Table 2. Occurrences of time expressions with *kun* in Finnish originals and Finnish translations (absolute frequencies in corpora of approx. 0.5 million words)

	Originals	Translations
silloin(kin/kaan) kun 'when'	90	127
heti kun 'as soon as'	64	100
sitten kun 'when'	59	56
aina kun 'whenever'	53	49
nyt kun 'now that'	33	72
juuri kun 'just when'	30	69
vasta kun 'not until'	29	26
sen jälkeen kun 'after'	27	71
samalla kun 'while'	18	32
sillä/samalla hetkellä kun 'at the same time as'	8	21
sillä välin kun 'while'	6	32
sillä aikaa kun 'while'	5	35
siitä asti kun 'ever since'	2	15
siitä saakka kun 'ever since'	2	6
siihen mennessä kun 'by the time'	2	15
Total	428	726
Total per 100 000 words	85.8	122.4

tion and negation, e.g. *ettei hän ollut*) show an interesting difference: *että* is more frequent in translations (1049.0 vs. 991.0), *ettei(vät)* in originals (138.5 vs. 106.8). Perhaps the lower frequency of *ettei(vät)* in translations can be explained by the nonexistence of a similar fused negative form in English. The SL conjuction and negation as discrete words may tend to trigger a similar structure in the TL. The other fused conjunctions *jottei, jollei, vaikkei* and *muttei* are very rare in both subcorpora and no considerable differences in their distributions were found.

The purpose conjunction *jotta* 'in order to' is surprisingly rare in Finnish originals in comparison with translations (11.2 vs. 56.7). The fact that *että* can be used synonymously with *jotta* fails to explain the difference, as *että* is also slightly less frequent in originals. However, *jotta* seems to often co-occur with *liian* + adjective 'too + adjective for' and *tarpeeksi* + adjective/noun 'adjective + enough to'/enough + noun to' in translations (22 occurrences in the entire translation subcorpus, 3.7 per 100 000 words) but not at all in originals, although both constructions are perfectly idiomatic. Examples (3) and (4) from the corpus are typical instances of such colligations.

(3) *Vuorokaudessa ei ollut **tarpeeksi tunteja, jotta** Cara olisi ehtinyt tehdä kaiken, mitä piti.*
literally: 'There weren't enough hours in a day for Cara to have time to do everything that had to be done.'

(4) *On **liian pimeää, jotta** hän voisi kävellä turvallisesti.*
'It's too dark for him to walk safely.'

On the basis of the children's subcorpora used in this study, these combinations with *jotta* can perhaps be considered translation specific colligations in the genre of children's literature, but this finding is not generalisable to other genres. Again, the translations are likely to reflect ST constructions.

4.2 Connectives more frequent in Finnish originals

The lower frequency of the relative pronoun *jossa* (sg.)/*joissa* (pl.) 'in which', 'where' in translations might be partly caused by the common use of corresponding participial attribute constructions, which was detected in the previous research (Puurtinen 1995). Another option which might occasionally be preferred to *jo(i)ssa* in translations when referring to location is *missä* 'where'. Indeed, *missä* is slightly more common in translations than in originals (26.3 vs. 17.4, including only those occurrences where *missä* and *jo(i)ssa* are both equally feasible alternatives). Perhaps the English *where* tends to get translated as *missä* rather than *jo(i)ssa*. Nevertheless, even the combined frequency of *missä* + *jo(i)ssa* turns out to be higher in Finnish originals (92.4 vs. 71.8) with no apparent reason.

Finally, the contexts and functions of the concessive conjunction *vaikka* '(al)though' show some interesting differences between originals and translations. One potential explanation for the somewhat lower frequency of *vaikka* in translations is choosing the longer construction *siitä huolimatta että* or *huolimatta siitä että* as a more direct equivalent for *despite the fact that*. However, this construction hardly occurs at all in either subcorpus (1.0 in originals, 1.2 in translations). Two verb forms used in connection with *vaikka* seem to be more common in translations: *vaikka* + verb + the clitic particle *-kin/kaan* (21.8 vs. 15.0) and *vaikka* + the conditional *-isi* (20.1 vs. 13.4). The only context, or meaning, of *vaikka* which is more typical of Finnish originals is 'on the other hand', 'but'; in other words, *vaikka* is not always a concessive conjunction but can also begin an afterthought of a kind to the previous clause and could be replaced with *tosin*, *kylläkin*, *mutta*, or *toisaalta*, as in the following examples from the subcorpus of Finnish originals.

(5)   *Lisko voisi olla kiva, **vaikka** en tiedä onko sekään erityisen seurallinen.*
      'A lizard might be nice, but I don't know if it's particularly sociable either.'

(6)   *Miksi hän ei ollut paremmin katsonut jalkoihinsa? **Vaikka** mitä se olisi auttanut?*
      'Why hadn't he watched his step more carefully? But what would that have helped?'

(7)   *Sinä olet vielä suurempi noita kuin minä aavistinkaan. **Vaikka** kyllä sinä aina olet erikoistapaus ollut.*
      'You are an even greater witch than I suspected. But you have always been a special case.'

Instead of the meaning 'despite the fact that', in these examples *vaikka* has the sense 'on second thought'. The frequency of such occurrences is 8.1 in translations (7% of all *vaikka* conjunctions) and 16.6, i.e. twice as high, in originals (13%).

5.   Conclusion

The above findings do not fully support the explicitation hypothesis, nor do they clearly contradict it. A few connectives are more frequent in translations, thus contributing to a higher degree of explicitation, while two connectives show an opposite trend with higher frequencies in originals. Thus, contrary to what might be assumed, high frequencies of NCs in translations do not seem to correlate with low frequencies of connectives. Instead of the frequencies, however, the most interesting findings are related to differences between the subcorpora in the contexts and functions of particular connectives. A more thorough analysis of the material is likely to yield additional information on such tendencies. Some of the differences can perhaps, unsurprisingly, be explained by ST features being reflected in translations, i.e. by a tendency to translate ST expressions literally. Other genres, such as academic literature or adult fiction, might reveal clearer patterns which distinguish originals and translations, as might also more homogeneous subcorpora of children's literature. The children's fiction included in the present corpus ranges from fairytales to girls' books and detective stories, and the age of the estimated readership from eight to twelve. In the same way as the overall style varies in different subgenres, explicitation may also show diverse patterns.

Notes

1. Explicit lexical realisations of clausal relations, as opposed to grammatical ones, are not discussed here (e.g. *syy epävarmuuteen* 'the reason for uncertainty').

2. In the previous study on the syntax of Finnish children's literature (Puurtinen 1995) no distinction was made between different types and functions of contracted clauses, and therefore no comparison between each type and the alternative structures with clause connectives is possible.

References

Baker, Mona (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target, 7*(2), 223–243.

Baker, Mona (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager* (pp. 175–186). Amsterdam/Philadelphia: John Benjamins.

Blum-Kulka, Shoshana (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17–35). Tübingen: Gunter Narr.

Chesterman, Andrew (this volume). Beyond the particular. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Daniels, Lucy (1996). *Koiranpentuja komerossa.* (Orig. *Puppies in the Pantry* 1994) Trans. Jaana Kapari. Helsinki: Tammi.

Daniels, Lucy (1998). *Mäyrä mummilassa.* (Orig. *Badger in the Basement* 1994) Trans. Jaana Kapari. Helsinki: Tammi.

Eskola, Sari (2002). *Syntetisoivat rakenteet käännössuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä* [Synthesising structures in translated Finnish. A corpus-based analysis of the special features of Finnish literary translations], [University of Joensuu Publications in the Humanities 30]. Joensuu: University of Joensuu.

Halliday, Michael A. K. (1994). *An Introduction to Functional Grammar.* 1985. London: Arnold.

Laviosa-Braithwaite, Sara (1996). Comparable corpora: Towards a corpus linguistic methodology for the empirical study of translation. In M. Thelen & B. Lewandowska-Tomaszczyk (Eds.), *Translation and Meaning, Part 3* (pp. 153–163). Maastricht: Hogeschool Maastricht.

Karvonen, Pirjo (1991). Kieliopillinen metafora ja sen vaikutukset tekstissä [Grammatical metaphor and its effects on a text]. In T. Lehtinen & S. Shore (Eds.), *Kieli, valta ja eriarvoisuus. Esitelmiä 18. kielitieteen päiviltä* (pp. 149–165). Helsinki: University of Helsinki.

Mauranen, Anna (2000). Strange strings in translated language: A study on corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects* (pp. 119–141). Manchester: St. Jerome.

Olohan, Maeve, & Baker, Mona (2000). Reporting *that* in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures, 1*(2), 141–158.

Øverås, Linn (1998). In search of the third code: An investigation of norms in literary translation. *Meta, 43*(4), 571–588.

Puurtinen, Tiina (1993). Kieliopilliset metaforat lastenkirjojen suomennoksissa [Grammatical metaphors in Finnish translations of children's books]. *Virittäjä, 4*, 546–563.

Puurtinen, Tiina (1995). *Linguistic Acceptability in Translated Children's Literature* [University of Joensuu Publications in the Humanities 15]. Joensuu: University of Joensuu.

Puurtinen, Tiina (2003). Nonfinite constructions in Finnish children's literature: Features of translationese contradicting translation universals? In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-Based Approaches to Contrastive Linguistics and Translation Studies* (pp. 141–154). Amsterdam: Rodopi.

Ravelli, Louise J. (1988). Grammatical metaphor: An initial analysis. In E. H. Steiner & R. Veltman (Eds.), *Pragmatics, Discourse and Text. Some Systemically-Inspired Approaches* (pp. 133–146). London: Pinter.

Scott, Mike (1998). *WordSmith Tools*. Oxford: Oxford University Press.

Vanderauwera, Ria (1985). *Dutch Novels Translated into English: The Transformation of a 'Minority' Literature*. Amsterdam: Rodopi.

Unique items – over- or under-represented in translated language?

Sonja Tirkkonen-Condit

University of Joensuu

One of the alleged universals of translation is the hypothesis that translations tend to over-represent linguistic features that are typical of the target language. On top of being counter-intuitive, this hypothesis seems to lack substantial empirical support. Among typical features are the linguistic phenomena that I call unique, i.e. linguistic items or elements which lack linguistic counterparts in the source language in question (see also Sari Eskola's article in this volume). The hypothesis of over-representation would predict that at least those unique items that are relatively frequent in a language should appear with a higher frequency in translated than originally produced language.

The hypothesis was tested by comparing the frequencies of two kinds of unique items in the Corpus of Translated Finnish, namely the verbs of sufficiency, such as *ehtii, mahtuu, jaksaa, malttaa* ('has enough time'/'is early or quick enough', 'is small enough', 'is strong enough', 'is patient enough', respectively), and the clitic pragmatic particles *-kin* and *-hAn*. The comparison shows that these uniquely Finnish items are *less frequent* in translated than original Finnish. It is suggested that the explanation for their under-representation in translated language should be sought in the translation process itself.

1. Introduction

Every language has linguistic elements that are unique in the sense that they lack straightforward linguistic counterparts in other languages. These elements may be lexical, phrasal, syntactic or textual, and they need not be in any sense untranslatable; they are simply not similarly manifested (e.g. lexicalized) in other languages. Since they are not similarly manifested in the source language, it is to be expected that they do not readily suggest themselves as translation

equivalents, as there is no obvious linguistic stimulus for them in the source text. Thus it might in fact be a universal tendency in translations to manifest smaller proportions of such language forms and functions which do not have similarly manifested linguistic counterparts in the source language. In other words, linguistic elements that are 'unique' in this sense would have lower frequencies in translated texts than in originally produced texts.

The frequency of unique items may affect the impression that a text makes on readers. I have some empirical ground to believe that the frequency of unique items influences the impression that the text makes on ordinary readers. A low frequency leads readers to think that the text is a translation, and a high frequency leads them to think that the text is original rather than translation. I carried out a test (see Tirkkonen-Condit 1998/2002) in which I asked native Finnish speakers to sort out a number of authentic text extracts into two piles: translated and original Finnish. When I analysed the two piles I noticed that the single linguistic phenomenon shared by those texts which most readers believed to be original texts – whether this was in fact the case or not – was their relatively high frequency of the unique elements. It is now possible to investigate the actual frequency of the unique items from the Corpus of Translated Finnish, which is a comparable corpus (see Mauranen 1998), and my purpose in this paper is to report on the results of this investigation.

2.   Purpose

The purpose of this paper is to test the Unique Items Hypothesis by checking the frequencies of some verbs and clitic particles using the Corpus of Translated Finnish which has been compiled at Savonlinna in a research project supervised by Professor Anna Mauranen. The verbs investigated here are verbs of sufficiency which constitute a lexical domain with no straightforward lexicalized translation equivalents in many Indo-European languages. These verbs have also attracted the attention of researchers of Finnish. Aili Flint's doctoral dissertation (Flint 1980) gives a semantic account of some forty such verbs.

The clitic particles investigated in the corpus are *-kin* and *-hAn*. The translation of the particle *-kin* depends on its pragmatic function, and in different contexts it translates differently, e.g. with the connectors *also, but, in contrast, consequently, thus*. The clitic particle *-hAn* is also multifunctional, and it usually conveys the assumption of shared knowledge along the same lines as the particle *you know* in spoken English (see Hakulinen 1976; Östman 1981, 1995).

The size of the corpus compiled in Savonlinna is now ten million words. The frequencies of the items in focus were checked from two genres, which I have labelled Academic and Fiction, each of which has a translated and original sub-corpus. Each of the four sub-corpora have about one million words, and the comparisons will be made between Original Fiction and Translated Fiction as well as between Original Academic and Translated Academic. Since there is every reason to believe that the genres Fiction and Academic are different in many respects, I will treat each genre separately. This means that my Original versus Translated comparisons will normally be done within each genre.

3.   Results

The quantitative results are presented in Tables 1 and 2 below.

The verbs are presented in Table 1 in the order of frequency in Original Fiction. Among the investigated verbs are the stylistically unmarked and relatively frequent verbs *ehtii*[1] ('has enough time', 'is early enough'), *jaksaa* ('is strong enough'), *uskaltaa* ('has enough courage'), *riittää* ('is enough'), *malttaa* ('is patient enough'), *viitsii* ('has enough initiative or interest'), and other, somewhat less frequent verbs from the semantic field of sufficiency.

The overall result of this investigation is such that it supports the Unique Items Hypothesis very strongly especially in the Fiction part: the frequencies are considerably lower in the Translated language corpus. Some verbs are almost entirely confined to fictional texts and hardly appeared at all in academic texts (e.g. *viitsii, kehtaa, viihtyy*). Thus the differences, if any, will not show clearly in a corpus of this size.

In addition to frequency comparisons, it is also interesting to compare the grammatical and collocational patterns that the verbs accumulate in Translated versus Original language. The verbs that do appear quite frequently in Fiction and Academic do not behave similarly in translated and original language. There are differences in their syntactic, semantic or collocational behaviour. For example, *uskaltaa* ('has enough courage or daring') has more instances of impersonal usage in Original Academic than in Translated Academic. *Viitsii* ('has enough initiative or interest') in Translated fiction is largely confined to the idiom *Älä viitsi*! ('Come on!'). *Malttaa* ('has enough patience') has a more varied use in Original Fiction than in Translated Fiction. In Original Fiction, for example, the following collocations are found: *malttoi mielensä* ('s/he controlled him/herself'), *malttaa olla tekemättä* ('s/he has enough control of him/herself not to…'), *malttaa odottaa* ('s/he is patient enough'). Translated

Table 1. Verbs of Sufficiency in Original vs. Translated Finnish Sub-corpora

		Fiction		Academic	
		Original (from English)	Translated	Original (from English)	Translated
		1,000,015	1,147,555	1,116,441	974,906
		Frequency per 1000 words	Frequency per 1000 words	Frequency per 1000 words	Frequency per 1000 words
ehtii	'has enough time'; is early/quick enough	0.499	0.324	0.094	0.026
jaksaa	'is strong enough'; 'has enough energy'	0.277	0.132	0.023	0.017
riittää	'is enough'	0.265	0.246	0.202	0.143
uskaltaa	'has enough courage; 'has the nerve to'; 'is brave enough'; 'is daring enough'	0.234	0.097	0.021	0.029
kelpaa	'is good enough'	0.096	0.045	0.032	0.004
mahtuu	'is small enough'	0.087	0.038	0.017	0.008
viitsii	'has enough initiative or interest'	0.080	0.096	0.004	0.005
kehtaa	'is bold enough'	0.069	0.012	0.009	0.001
viihtyy	'is comfortable enough'	0.064	0.039	0.004	0.008
malttaa	'is patient enough'	0.050	0.020	0.004	0.002
rohkenee	'is brave enough'; 'has enough courage'	0.037	0.009	0.005	0.007
joutaa	'is idle enough'	0.020	0.007	0.001	0.000

Fiction, in contrast, is largely confined to [*tuskin*] *malttoi odottaa* ('could [barely] wait to…'). *Riittää* in Translated Academic is largely confined to the syntactic construction *riittää* plus third infinitive, such as *riittää osoittamaan, selittämään, kuvaamaan, korostamaan, perustelelmaan, opettamaan* ('is enough to show/explain/describe/emphasise/justify/demonstrate'), whereas Original Academic has a wider range of syntactic constructions.

Table 2 shows that the clitic particle *-kin* is a frequent phenomenon in Finnish. In each of the sub-corpora of 950,000 words, the particle has roughly 5000 to 7000 appearances. It is slightly more frequent in Academic than in Fictional texts, and it is systematically more frequent in Original Finnish than in Translated Finnish. There are about 7 instances per one thousand words in

Table 2. Particles *-kin* and *-hAn* in Original vs. Translated Finnish Sub-corpora

	Fiction				Academic			
	Original 950,000 words		Translated 950,000 words		Original 950,000 words		Translated 950,000 words	
	Total	Per 1000 words	Total	Per 1000 words	Total	Per 1000 words	Total	Per 1000 words
-kin	6595	6.942	4810	5.063	6895	7.258	5579	5.873
-hAn	1856	1.954	1216	1.280	635	0.668	251	0.264

Original Fiction versus 5 instances in Translated Fiction, and 7 instances per one thousand in Original Academic versus 6 instances in Translated Academic. As was noticed in the discussion on the verbs above, the difference between Original and Translated is again more marked in Fiction than in Academic.

The clitic particle *-hAn* is less frequent than *-kin*, but it is frequent enough to warrant a comparison for the purposes of the Unique Items hypothesis. Since *-hAn* has a more frequent use in colloquial language, its greater frequency in Fiction was to be expected. It has about 2 appearances per one thousand words in Original Fiction, as against 1 in Translated Fiction. In Original Academic it has about 0.7 appearances per one thousand words as against 0.3 in Translated Academic

The research on clitics supports the Unique Items Hypothesis very strongly. The clitics provide an even better testing platform for the hypothesis than lexical items, since they are stylistically relatively unmarked. Moreover, in translation from Finnish into English or German, for example, the clitics present a decision point for the translator. As they lack straightforward linguistic counterparts, they call for a semantic and pragmatic analysis in each context. In translation from English or German into Finnish, on the other hand, the source texts do not display any items that *need* to be translated by these clitics. The items that *could* be translated by clitics are also translatable by lexicalized connectors.

4.   Discussion

The most obvious explanation for the relative scarcity of the verbs of sufficiency in Translated language is the explanation suggested by the Unique Items Hypothesis itself, namely that translators dismiss these verbs because they are not obvious equivalents for any particular items in the source text. The verbs

do not therefore suggest themselves as first choices for translators, even where they would fit the context very well. When the *has enough/is enough* pattern appears in an English source text, the translator is led to imitate the pattern and to generate *on tarpeeksi* instead of one of the above verbs. The clitics, too, can easily be dismissed in translation, since *thus* translates by *niinpä*, *also* by *myös*, *but* by *mutta*, etc.

Translation scholars have noted this type of source language dependence before, but it has not been studied systematically in corpora. Katharina Reiss pointed out the problem of "missing words" in her book on translation quality assessment (Reiss 1971). She suspected that translators did not perhaps fully exploit the linguistic resources of the target language. As one of the devices for translation criticism to be used without recourse to the source text, Reiss suggests in line with Güttinger (1963: 219), that you carry out a simple test. Take the most frequent words in the target language that do not exist in the source language and check the extent to which these appear in the translation. These "missing words" will reveal whether the translator knows the target language well enough to attain good translation quality. This rule of thumb, according to Reiss (1971: 19), applies not only to the "missing words", but to "alle Begriffe und Wendungen, die in der anderen Sprache mit unterschiedlichen sprachlichen Mitteln zum Ausdruck gebracht werden".

Miriam Shlesinger (1992) noticed in student translators a failure to lexicalize – to use one word instead of many – in instances where the source language expressed the idea with several words, whereas the target language would have called for a single word equivalent. This tendency was noticed in professional translators, too, and not only in students. Thus the English words *deadline* and *shortlist* did not readily find their way even to the English translators' texts who translated from Hebrew into their mother tongue.

Gideon Toury (1995: 224–225) suggests that translation as a process causes a tendency to resort to expressions that bear resemblance to the SL rather than expressions that are typical in a similar context in the target language. Toury (1995: 225) suggests that "this is highly indicative of the fact that the requirement to communicate in translated utterances may impose behavioural patterns of its own." The literal equivalents and attempts to translate word by word are very frequent in think aloud protocols of translation even in the performance of translators whose target texts do not show a tendency to translate word for word. The literal expressions may be used as a way to 'listen to' what the expression means prior to venturing a translation proper. If the literal equivalent makes perfect sense and does not violate the target language norms, there is no immediate reason to discard it. This is why the

unique elements tend to be less frequent also in published translations than in original writing.

5.   Conclusion

The reasons why the unique linguistic phenomena tend to be under-represented in translated language may be found in a (potentially universal) tendency of the translating process to proceed literally to a certain extent. This means that the translator picks out lexical items, syntactic patterns and idiomatic expressions from his bilingual mental dictionary, and this is what happens. The *has enough* pattern tends to generate the *on tarpeeksi* pattern and the connectors *also*, *thus*, and others tend to generate connectors *myös*, *niinpä* etc. Since the verbs and the particles discussed here do not have linguistic counterparts, they do not appear in the bilingual mental dictionary and there is nothing in the source text that would trigger them off as immediate equivalents. Thus they have a slighter chance to be chosen into the target text than they have of appearing in texts that are produced in the original.

Note

1. The verbs are introduced in their third person singular forms. Thus *ehtii* can be glossed as 'has enough time to…' or 'is early enough'.

References

Eskola, Sari. (this volume). Untypical Frequencies in Translated Language: A Corpus-Based Study on a Literary Corpus of Translated and Non-Translated Finnish. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Flint, Aili (1980). *Semantic Structure in the Finnish Lexicon: Verbs of Possibility and Sufficiency*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Güttinger, Fritz (1963). *Zielsprache: Theorie und Technik des Übersetzens*. Zürich: Manesse Verlag.

Hakulinen, Auli (1976). Liitepartikkelin -han/-hän syntaksia ja pragmatiikkaa [On the syntax and pragmatics of the clitic particle -hAn]. *Meddelanden från Åbo Akademi Forskningsinstitut, 7*, 1–60.

Östman, Jan-Ola (1981). *'You know': A discourse-functional approach* [Pragmatics and Beyond II: 7]. Amsterdam: John Benjamins.

Östman, Jan-Ola (1995). Pragmatic particles twenty years after. In B. Wårvik, S. K. Tanskanen, & R. Hiltunen (Eds.), *Organization in Discourse. Proceedings from the Turku Conference* (pp. 95–108). Turku: University of Turku.

Mauranen, Anna (1998). *Käännössuomi ja kääntämisen universaalit: tutkimus korpusaineistolla* [Translated Finnish and Translation Universals: A Corpus Study]. Project research plan, MS. Savonlinna School of Translation Studies.

Reiss, Katharina (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. Munich: Max Hueber.

Shlesinger, Miriam (1992). Lexicalisation in translation: An empirical study of students' progress. In C. Dollerup & A. Loddegaard (Eds.), *Teaching Translation and Interpreting. Training, Talent and Experience* (pp. 123–127). Amsterdam/Philadelphia: John Benjamins.

Tirkkonen-Condit, Sonja (1998/2002). Translationese, a myth or an empirical fact? A study into the linguistic identifiability of translated language. Paper read at The 24th University of Wisconsin-Milwaukee Linguistics Symposium, September 10–12, 1998. *Target, 14*(2), 207–220.

Toury, Gideon (1995). *Descriptive Translation Studies and beyond*. Amsterdam/Philadelphia: John Benjamins.

Universals in the translation class

What happens to "unique items" in learners' translations?

"Theories" and "concepts" as a challenge for novices' views on "good translation"

Pekka Kujamäki
University of Joensuu

This article reports on an experiment, in which two components of translator students' professional self-understanding were challenged: their doubts on the relevance of theoretical knowledge as part of their professional competence as well as the strong belief in their L1 competence. The experiment draws on Toury's "law of interference" and the analysis of students' translations is based on Tirkkonen-Condit's hypothesis that unique, TL-specific elements may be underrepresented in translations. Students' translations of short English and German source texts are consistent with this hypothesis: the experiment reveals that even a source text that seems to present no translation difficulties in surface structure is still a powerful constraint in translation and produces language patterns which are alien to or at least deviant from non-translated target language usage as revealed in this experiment by a small-scale cloze test.

1. Introduction

From the very beginning, students of translation seem to have a strong but biased understanding of the essential components of their competence and how to develop them. One common impression that manages to survive despite the challenges presented by teaching is their suspicious view of the role of theoretical knowledge as an essential part of their studies as well as of their competence. "Theory is theory and practice is practice", is the argument that a teacher of e.g. research seminars is regularly confronted with. Theorising is seen as a self-sufficient activity, linked to translator students' lives only to make it difficult and to take learners' time from practical translation exercises. And

we know that this line of thinking is common among professional translators as well (see e.g. Chesterman 1993b; Hönig 1995:25, 158). However, if we dig deeper into this view – e.g. by eliciting explicit written commentaries – it soon turns out that students' frustration often comes from their experiences in the translation class, where, in the name of practice, "theories" and "concepts" are still too often kept out of everyday business.

Another common feature of translator students' (semi)professional self-understanding is the unfaltering belief in their L1 (in this case: *Finnish*) competence when they translate from the FL. In L1 translation students seem to regard comprehension of the FL source text as the main problem of the task, so that after having understood the text – and taking into account the purpose of the translation – the formulation of an adequate and natural target language text should be no problem. One interesting expression of this (perhaps learned) faith was a discussion on Toury's "law of interference" (1995: 275) with my 3rd year seminar students in spring 2001: it turned out that in this era of functionally oriented translation, learners do not (dare to) regard "translationese" or "interference" as relevant topics of research on the (most certainly learned) argument that "these phenomena *should not* exist anyhow". In face of the evidence provided by descriptive research on translation so far (e.g. by Toury himself 1995:206–220) this reasoning sounds rather odd. But when compared with the line of argumentation for example in Schmidt (1989), where "interference" is defined either as avoidable deviations from correct target language usage or as insufficient and incomplete reception of the source text (Holz-Mänttäri 1989:132), such commentaries make, after all, perfect sense (for a broader view on interference see e.g. Mauranen, this volume, Eskola, this volume).

These two observations provoked me to carry out a small experiment. The idea was, on one hand, to question this self-confidence and show students that even a source text that seems to present no translation difficulties in surface structure (e.g. in the form of potential "false friends") is still a powerful constraint in translation and very likely to produce language patterns in translation which are alien to or deviant from general target-language usage. For this purpose – as well as to argue, on the other hand, for the applicability of at least some "concepts" and "theories" in classroom practice – the experiment was based on Sonja Tirkkonen-Condit's hypotheses (2000, 2002: 16 and this volume) that TL-specific elements, "unique items", may be underrepresented in translations. I created a short text that deals with driving in Finland in winter and includes several "unique" nouns, which refer to snow or Finnish weather conditions. The text was translated into German and English by

native speakers, and finally back-translated into Finnish by 36 students in Savonlinna as well as in Tampere. At the next phase the translated items were compared with students' non-translated language use as revealed by a small-scale cloze test.

This article discusses the results and some implications of the experiment. The experiment took place in the context of translation exercises, which usually involve some kind of normative statements and value judgements (Chesterman 1993a). Nevertheless, in this article the approach is mainly descriptive, the pedagogic goal is to make learners aware of what they are doing by identifying at least some features of their task performance, to show their consequences, and to challenge students' vague views on what translation is all about.

2.   Unique items

The concept of "unique items" has been recently introduced by Tirkkonen-Condit, who suggests that it might be a universal tendency of translated language "to manifest smaller proportions of such language forms and functions which do not have *straightforward* equivalents in other languages" and, in particular, in the source language in question (Tirkkonen-Condit 2000, my emphasis; see also Tirkkonen-Condit this volume). "Unique items" can be seen as a rather broad and dynamic category of linguistic features which covers lexical, phrasal, syntactic or textual elements (Tirkkonen-Condit 2000) and even pragmatic functions (Mauranen 2001) and whose extension is usually different from one language pair to another. The emphasis on "straightforward" refers to the fact that these elements or functions need not be untranslatable at all:

> Rather, very often they seem to have only partially overlapping equivalents in other languages, i.e. equivalents *that tend to explicate* their implicit SL meaning, which is part of the world knowledge of the SL user.
>
> (Tirkkonen-Condit 2000, my italics)

In other words, lexical items such as the Finnish expressions for "snow" (*hanki, kinos, nuoska* etc.) very often carry semantic-pragmatic distinctions that are usually not habitual or necessary in any other languages. Or they may be items that are semantically ambiguous in the sense that they can be used in different pragmatic situations, in which the L1 speaker usually knows automatically what is meant by the word. For instance the Finnish word *keli* ('surface conditions') can be used in different contexts with reference to driving conditions on the road, to skiing conditions in the woods or even to sailing

weather out on the sea. The ambiguity can also be seen in translations of these items into other languages (see Appendix 1) and in their dictionary equivalents. In both cases the used or the potential translation equivalent is either much more explicit than *or* only a semantic approximation of the Finnish unique item in question. The following table (Table 1) with a few examples from Finnish-German and Finnish-English dictionaries illustrates the tendency of dictionary equivalents to explicate the meaning of such items in other languages. It also shows the close (both semantic and functional) synonymy of the first two Finnish items.

Table 1. Dictionary equivalents of *hanki*, *kinos* and *keli* in *Finnisch-Deutsches Grosswörterbuch* (Katara & Schellbach-Kopra 1997 = FD) and *Finnish-English General Dictionary* (Hurme, Malin, & Syväoja 1984 = FE)

Hanki	Schneefläche, Schneedecke, Schneekruste, Schnee (FD)	→ *lumikerros, lumikasa, lumi*
	snow; (kinokset) snowdrifts (FE)	→ *lumi, lumikasat*
Kinos	Schneewehe, -verwehung (FD)	→ *lumikasa*
	drift, snowdrift, snowbank (FE)	→ *penkka, lumikasa, lumipenkka*
keli	Zustand der Wege (im Winter), Straßenzustand; (säästä) Wetter, Witterung (FD)	→ *teiden kunto, tieolosuhteet, sää*
	conditions, road/snow/surface conditions, weather (FE)	→ *olosuhteet, tieolosuhteet /lumiolosuhteet/tien pinta, sää*

To sum up: the category of "unique items" in a sense gets a definition in translations from L1 to L2. As a research object, however, the category is interesting in translation into the opposite direction, i.e. from L2 to L1. As implied by the above examples, the concept of "unique items" opens (not always, but with lexical units like these) a new perspective into the problem of realia in translation: instead of asking the old-established question of how these culture-specific items of L1 could be or have been translated into other languages (L2), the question to be asked now is, whether and how such realia of the target culture are used in translated utterances. Since "unique items" like the ones above are lexicalised in Finnish but not in the source languages from which the translation *into* Finnish is (e.g. in this particular experiment) taking place, i.e. from German and English, the source languages do not offer any direct stimulus for their use. The interesting research question is then, what happens to such lexical elements in learners' translations into Finnish. Are they represented at all? With respect to students' belief in their L1 competence one would expect a straigtforward positive answer. To be more realistic, however,

Tirkkonen-Condit's above hypotheses and her results from the comparable Corpus of Translated Finnish (CTF) lead us rather to assume that in learners' translations unique items of Finnish are used less than so called lexical or word-for-word translations that are stimulated by the English or German source text surface structure (as seen in the right column of Table 1).

3.   Design of the translation test

The source texts used in the translation experiment were themselves translations produced by native speakers of German and English respectively from a Finnish text written by the present writer. This arrangement was necessary for the following reasons:

In experimentation, the difficulty of the source text is one potential variable among many others in translator performance. As Riitta Jääskeläinen (1999: 245) points out, it is "likely to influence the number of problems and the choice of appropriate strategies, but also the subjects' ability and/or willingness" to perform the task according to given specifications. Students' translation performance, be it in experimentation or in "normal" classroom practice, is very vulnerable to source text difficulty. With difficult texts the risk of frustration and tiredness is high, as a great deal of students' effort during the translation task can be taken up by extensive source text processing alone (see Jääskeläinen 1999: 198; Jääskeläinen & Tirkkonen-Condit 1991). For the purposes of this experiment, in which target language rendering was in focus, I needed a text which allowed students to concentrate on target text production instead of investing *too much* energy in understanding and analysing the source text. Consequently, I needed a source text that would put the students in a thematic context that was an inherent part of their world knowledge.

To make this experiment as convenient as possible for students (i.e. to have as short a translation task as possible) and for myself (to avoid a weary search in newspapers, magazines, the Internet etc. for a short text that could be used in an experiment of this kind), I created a text of my own dealing with *driving a car in Finland in winter*. In this Finnish text I inserted the above mentioned more or less culture-specific realia *keli, kinos* and *hanki*. This Finnish text was then translated into English and German by my native speaker colleagues (see Appendix 1).[1] Their commission was to regard the source text as a first part of a longer text which was to be translated and published on the web site of the National Motoring Organisation in Finland (http://www.autoliitto.fi/eng.cfm

and link "Motoring in Finland"). These texts, then, were given to students as new source texts, with their origin as translated texts hidden. Their assignment was specified as follows: the text (sample) is to be translated for the magazine *Etumatkaa* – a quarterly customer magazine of the Volkswagen group in Finland. Its winter issue plans to publish articles where foreigners write on their driving experiences in Finland in winter. In Appendix 2, the scanned excerpt of the customer leaflet posted by the local Volkswagen dealer in autumn 2001 reveals the thematic relevance of such translation tasks and provides one authentic text example of the above discussed item *keli*.

The German text "Winter, oh weh!" was translated in Savonlinna by students of my 3rd year proseminar group in April 2001 (test group A; N = 13), as part of my 2nd year course Translation German-Finnish in September 2001 (group B; N = 10), and additionally in Tampere by a couple of 2nd and 4th year students of German Translation and Interpretation (group C; N = 6). Finally, the translation of the English text "An Ordinary Winter's Tale" was provided by seven Savonlinna students of the English 3rd year proseminar group in September 2001 (group D), bringing the total to 36 translations.[2] The students were allowed to do their translation when and where they chose (alone, however, *and* within the limits of a rather loose deadline) and to consult the reference material they normally use.

Before we take a closer look at the students' solutions, a caveat may be in order. Since the source texts are translations of a fabricated Finnish-language text, no normative comparisons between the Finnish original and the learners' translations are to be made. The way I look into the translational solutions with a specific interest in selected culture-specific realia *does not imply that only the use of these realia in target texts equals a correct translation and excludes the others*. Rather, the main interest is – in the class-room situation as well as in the later steps of this experiment – didactic: what do the translations as products possibly tell learners about their own processes?

4.   "Unique items" in learners' translations?

The results of the translation test are shown in Tables 2, 3 and 4. The summaries show translational patterns in favour of solutions that are directly motivated by the lexical surface structure of the source texts in question. The tendency to overlook rather than to use realia such as *hanki*, *keli* and *kinos* is clear and as such consistent with the above hypothesis concerning unique items in translation into L1. In each case all test groups show a very similar distribution:

Table 2. Learners' translations of *Straßenverhältnisse* and *conditions* into Finnish

"Keli"	*die Straßenverhältnisse* wurden immer miserabler	N = 29
a.	*tieolosuhteet* (12), *ajo-olosuhteet* (4), *liikenneolosuhteet* (1)	17
b.	*katujen kunto [huononi]* (1), *teiden ajokunto* (1), *tiet [surkeammaksi/huononivat]*, (2)	4
c.	**keliolosuhteet** (3), **ajokeli** (2)	5
d.	**keli**	3
	conditions rapidly worsened	N = 7
a.	*olosuhteet* (2), ajo-olosuhteet (1), sääolot (1)	4
c.	**ajokeli**	1
d.	**keli**	2

Table 2 shows that five translations out of 36 manifest the item *keli* as a translation for *die Strassenverhältnisse* or *conditions*, in further six solutions the item is part of the compounds *ajokeli* ('driving' + 'conditions') and *keliolosuhteet*: according to a contemporary monolingual dictionary of Finnish (*Suomen kielen perussanakirja*, Haarala et al. 1990–1994) the former is a contextual and more explicit synonym of *keli*, whereas the latter presents semantic tautology ('conditions' + 'conditions') and thus a very explicit wording of the concept. The rest include mainly lexical translations (*tieolosuhteet* 'road conditions') or semantic approximations of the German or English source text item.

Furthermore, as shown in Tables 3 and 4 overleaf, the unique item *kinos* (or its verb form *kinostaa*) is used in eight translations (each time in target texts that do not contain the noun *keli*), whereas the item *hanki* is not used at all. As Table 4 reveals, later in the text two more students came up with *kinos*, an example of the synonymy of these two culture-specific items. In both cases the overwhelming majority, however, seems to favour close lexical translations of the German or English surface structure that retain the source texts' explicit wording of the concept.

Analogous to the above mentioned *ajokeli*, the compounds *lumikinos* and *lumihanki* are interesting examples of this: they reveal students' awareness of the Finnish unique items, but at the same time they seem to be careful to ensure that the source texts' explicitly expressed semantic component "snow" (*lumi*) is manifested in the translations as well.

Table 3. Learners' translations of *den Schnee zu Häufchen häufeln* and *snowbank* into Finnish

"Kinos"	...häufelte **den Schnee** am Straßenrand **zu** schon recht ansehnlichen **Häufchen** auf.	N = 29
	lunta – töyräiksi (1), *kasoiksi/kasoihin* (10), *vall(e)iksi* (3), *keoiksi* (1), *penkoiksi* (2), *tienreunaan* (1)	18
	lumikasat	2
	lumikinoksiksi	2
	lunta...kinoksiksi	7
	...left a low **snowbank** on the side of the road.	N = 7
	lumipenkka	3
	lumikinos	3
	kinosti lunta	1

Table 4. Learners' translations of *mitten im Schnee* and *snowdrift* into Finnish

"Hanki"	...fand ich mich in meinem Wagen **mitten im Schnee** wieder.	N = 29
	lumen – keskellä (13), *saartamana* (1), *ympäröimänä* (1)	15
	keskellä – lumikasaa (3), *lumipenkkaa* (4)	7
	keskellä lumisohjoa	1
	keskellä **lumikinosta**	3
	kinoksesta	2
	keskellä **lumihankea**	1
	...found myself and my car stuck in a **snowdrift**.	N = 7
	juuttuneena	
	– lumipenkereeseen (1), *lumipenkkaan* (1)	2
	*– **lumikinoksessa***	2
	*– **kinoksessa***	2
	*– **lumihangessa***	1

5. First explanations

To explain observations of this kind Tirkkonen-Condit (2000) refers to the explanation suggested by the hypotheses about "unique items" itself. Adapted to this study it reads as follows: since the items are not lexicalized in the source languages from which the student translation is taking place here, they do not suggest themselves as obvious first-choice-equivalents for the source text expressions. Rather, the German and English stimuli seem to suggest into the target text straightforward lexical or dictionary translations. On the basis of

evidence from corpus-based as well as from process research on translation, Tirkkonen-Condit suspects in her recent article a

> filtering element in the translation process which directs the translator's mind to those linguistic elements in the target language that *do* have linguistic counterparts. This filter blinds the translator so that s/he tends to overlook the unique linguistic items. (Tirkkonen-Condit 2002: 16)

As Tirkkonen-Condit (2000) notes, if the filtered literal equivalents make perfect sense and do not (seem to) violate the target language norms, translators may find no immediate reason to give them a second thought. As a consequence, unique items (such as *hanki, keli* and *kinos* in this experiment) are used only occasionally and – as Tirkkonen-Condit's research on "verbs of sufficiency" and Finnish clitic pragmatic particles (ibid.) from the comparable corpus of translated and non-translated Finnish indicates – less frequently in translated than in non-translated Finnish.

However, the present small-scale translation test does not allow comparisons or generalizations of this kind yet, as we have no information about the learners' non-translated utterances. The translation test leaves one essential question unanswered, namely, to what extent the learners really actively use or passively know the studied realia *keli, hanki* and *kinos*.

6. A control test

To answer this question, a small control test was created in order to compile (an imitation of) a comparable corpus of students' language use in translated and non-translated utterances. The control test involves a mixture of a "cloze test", inspired by an empirical test conducted by Mary Snell-Hornby (1983; see also Vannerem and Snell-Hornby 1986), and the method of "picking out scene elements in a frame" introduced by Paul Kussmaul (see e.g. 2000a, 2000b) to enhance the processes of creative translation in classroom situations. Both Snell-Hornby's and Kussmaul's ideas are based on the cognitive model of scenes-and-frames by Fillmore (1977). Snell-Hornby uses a "mini-cloze" (Toury 1991: 48) and a visual presentation of the same text to collect spontaneous supplements for one missing simile. In other words, Snell-Hornby is looking for habitualized linguistic choices, *frames*, that get regularly associated with a certain mental, in this case verbalized or visualized picture, i.e. a *scene*. These choices are subsequently compared with the students' translations of the same text and the same simile. In Kussmaul's method, in turn, the meaning of a specific word is dis-

cussed by picking out details of the scene that belong to this frame, and in this way students are helped to find more creative translational replacements for difficult source text frames.

In the present case, "picking out scene elements" was conducted with a story similar to the above mentioned text in the translation test, which was, however, told freely by the present experimenter in front of the class. The idea was to play a helpless translator, who had a minor translation problem in the form of three missing Finnish words that would make the text complete. The students' task was, then, to picture with the given elements of the story this specific scene and write down for each of the three gaps those (max. 3) lexical candidates that they first came up with and that seemed to fit in the particular, described context. The cloze test was conducted in Savonlinna with three groups, in total 38 first or second year students (13 + 12 students of English translation and 13 students of German translation).[3] The results of the control test are presented in the following two tables:

Table 5. Students' proposals for the missing word, cloze item "keli"

"Keli"	First choice	Second choice	Third choice	Total / N = 38:
keli	26	5	2	33
ilma, sää ('weather')	6	10	6	22
others: *tuuri, mäihä* ('luck'), weather	6	4	1	11

As can be seen in Table 5, as many as 33 students out of 38 suggested the word *keli* for the first "problematic" gap, and 26 of them gave it as their first candidate. On this basis it seems justified to maintain that these words still are habitualized nouns, even in students' language use, when it is not constrained by foreign language stimuli. Other, clearly less frequent candidates include situationally adequate expressions such as "weather" and expressions like "luck", which reveal a slightly different, though expected and acceptable interpretation of the scene. One student misunderstood the task altogether and proposed consequently three English words. Additionally, it is interesting to observe that in their responses the students never used the expressions *ajo-olosuhteet, liikenneolosuhteet* and *tieolosuhteet* that were, in contrast, frequent in the translated texts.

In the latter two cases of *kinos* and *hanki*, as seen in Table 6, the variation is already wider but the unique items in question are nevertheless still more frequently used in this non-translational situation than in the translation test: more than one third of the students use the words *kinos* and *hanki*. The

Table 6. Students' proposals for the missing words, clozes "kinos" and "hanki"

"kinos"	First choice	Second choice	Third choice	Total/N = 38:
kinos	**11**	1	–	**12**
lumikinos	**3**			**3**
lumikasa, lumikeko, lumivalli, lumipenkka ('snowdrift'/'snow-bank')	11	2	1	14
keko, valli, kasa, penkka ('drift' / 'bank')	11	4	2	17
others: *ura* ('trail'), *weather*	2	2	–	4
"hanki"				
hanki	**11**	–	–	**11**
lumihanki	**2**			**2**
<u>*kinos*</u>	6	1	–	7
lumipenkka, lumikasa, lumivalli	3	1	–	4
lumi ('snow')	6			6
penkka ("bank")	2			2
others: *loska, sohjo* ('slush'), *kiinni* ('stuck'), *pelto* ('field'), *ditch*	6	3		9
"no answer"	2			2

difference between the translated and non-translated language use becomes clearer, if the frequent use of *kinos* as an adequate situational synonym for *hanki* is taken into consideration.

Furthermore, it is interesting to observe that in this non-translational cloze test the slightly tautological synonyms *lumikinos* and *lumihanki* as well as other compounds are less frequent than in students' translations. In addition, cloze test wordings include situational equivalents like *keko, penkka* and *valli* ('drift' or 'bank') as well as such realia as *loska* and *sohjo* ('slush'), which both refer to the element of "snow" in this particular scene, and are practically not at all manifested in translations.

7. Concluding remarks

Together these observations on novice translators' translated and non-trans-lated language use are but one example of the functions of the "law of

interference" (Toury 1995:275), i.e. the hypotheses about the influence of the source text surface structure on translated target-language use. Conditioned by the source text, learners' translation processes produce a distinct distribution of lexical choices that give the target text "a taste of translationese" (Tirkkonen-Condit 2002:12) and in any case make the text semantically more explicit than their non-translated expressions. As such these results comply with earlier findings on features of translated language, whether obtained from empirical tests on novices or professionals (e.g. Snell-Hornby 1983) or from research on larger corpora of authentic translated or non-translated texts (e.g. Olohan & Baker 2000; Eskola, this volume; Tirkkonen-Condit 2000 and this volume).

It is therefore convenient to sum up the results of this experiment with Toury's comment on the data provided by Snell-Hornby's above mentioned experiment:

> It would seem, then, that even people who are well aware of so-called native "situational equivalents", and use them in comparable native-like situations, tend to ignore them as translational replacements, *even if they are trained to try and establish translation relationships on the highest possible level (as the subjects of this experiment, being students of translation in a modern institute, definitely were)*. To me this is highly indicative of the fact that the very need to "communicate in translated utterances" (Toury 1980) imposes patterns of its own, a statement which certainly deserves some more consideration – and specification. In experimental methods too.   (Toury 1991:50, my italics)

Toury's conclusion is easy to agree with. With respect to classroom practice I would like to add, as implied by the added italics above, that such observations are also pedagogically relevant: is it really the case that our students try to establish translation relationships on the highest possible level?

If we take, once more, a look at the students' translations, it is easy to see that the translations that did not use, for example, the Finnish realia *keli* share one feature, namely the semantic component "condition", which was mani-fested in the source texts either as (-)*verhältnisse* or as *conditions*. As pointed out in the beginning, it is a semantic component that is not expressed in the Finnish word. All in all, the students' target texts imply an adherence to a concept of translation that involves an understanding and rendering of words or, at best, of sentences rather than texts let alone *scenes* behind the source text's linguistic surface. Seen from the perspective of the control test, it seems that students are unable or reluctant to "dive" into the context and exploit it for reconstructing the situation and for releasing themselves of the SL-surface structure to fully construct the scene or the mental model involved in the text. Hence students do not find natural TL frames for the given scene. In research

on translation processes, this phenomenon has been described in several ways, depending on the perspective (and experimental conditions, subject population and the data), e.g. "form-oriented learner translation" vs. "sense-oriented professional translation" (Lörscher 1987), "shallow processing" as a feature of non-professional and unsuccesful behaviour and semi-professional translators' growing awareness of potential translation problems (Jääskeläinen 1999:202; Tirkkonen-Condit 1987) or "unmonitored equivalence generation" manifested in novices' performance (Tirkkonen-Condit 2002:12). The basic observation, however, remains the same: the idea or ideal of establishing translation relationships on higher levels than that of one single word or compound has not been adopted by (most of the) volunteers of this experiment, yet.[4]

Such task performance may have to do with a number of so called *implied rules of translation* in novice translators' thinking. As Hönig (1988:158, 1995:25) has shown, these include rules such as "translate word-for-word whenever possible and as freely as necessary", "translate as exactly as possible", so that "the correctness of the translation" can be checked with bilingual dictionaries. According to these rules translation is inevitably poorer than the source text and usually sounds odd or in any case not like a non-translated target language original. This is, as the train of argument goes, inevitable and therefore quite normal (ibid.). The repertoire of such and other encultured rules not only defines the way the discourse on translating, translations and – consequently – on the status and role of translators is manifested in our contemporary society, be it by users of translators' services, in reviews of translated novels, in foreign language exercises in schools and universities, in layman discussions on the correctness of subtitles etc., but also constitutes the basis of students' "*translatorische Inkompetenz*" (Hönig 1988: 156) at the beginning of their studies. Moreover, it will continue to mark their translation performance, if our teaching is unable to challenge this disposition. After all, in an endeavour to construct a more realistic view of translation processes and translations as products, of features of expertise and/or professionality etc., scientific knowledge provides an evident tool kit. This is why "theories", "models", "concepts" and experimentation with them should have an essential role in the pedagogics of translation, not only in research seminars but also and above all in the translation class: they open a way to novices' better understanding of their future status as experts of human translation.

Notes

1. I thank Stephen Condit and Martina Natunen, Savonlinna School of Translation Studies, for their translations.

2. I am very grateful to Riitta Jääskeläinen, Savonlinna School of Translation Studies, and Dieter Hermann Schmitz, University of Tampere, School of Modern Languages and Translation Studies, for their assistance.

3. I thank Riitta Jääskeläinen and Unto Sinkkonen, Savonlinna School of Translation Studies, for their assistance.

4. This hypothesis will be tested at the next stage of this project, where students' translation processes are recorded for analysis with screen recording software (see Kujamäki 2003).

Appendix 1: The source texts of the present experiment

SE TAVALLINEN TALVINEN TARINA

Tänään aamulla herättyäni huomasin kauhukseni, että yöllä oli satanut syksyn ensimmäinen lumi. Eikä minulla ole talvirenkaita autossa tai edes autotallissa! Aamun linja-auto kaupunkiin oli kuitenkin ehtinyt jo mennä, joten minun oli pakko lähteä ajelemaan töihin omalla autolla. Rengasliikkeestä saisin sitten turvalliset renkaat alleni.

Matkalla tuiskuava lumi muuttui rännäksi ja **keli** vain paheni. Edellä ajelevan aura-auton jäljiltä tien viereen jäi jo matalia **kinoksia**. Ajelin hiljaa mutta en sitten kai kuitenkaan tarpeeksi varovasti: yhdessä risteyksessä auto vain lähti puskemaan kohti tien oikeaa reunaa, ja pian löysin itseni ja autoni **hangesta**. Hieno alku päivälle!

Vaikkei itselleni kummemmin käynytkään, tiesin kuitenkin siinä apua odotellessani, millaisin otsikoin tämän aamupäivän liikenteestä raportoitaisiin huomisen lehdessä: (…)(112 words)

WINTER, O WEH!

Als ich heute morgen aufwachte, bemerkte ich zu meinem Entsetzen, dass es in der Nacht zum ersten Mal in diesem Herbst geschneit hatte. Und ich hatte noch nicht die Reifen gewechselt, besaß nicht mal Winterreifen! Den morgendlichen Bus in die Stadt hatte ich schon verpasst, so musste ich wohl oder übel mit dem Auto zur Arbeit fahren. Beim Reifenhändler würde ich mir dann Winterreifen montieren lassen.

Unterwegs verwandelte sich der Schnee in Schneeregen und die Straßen-verhältnisse wurden immer miserabler. Vor mir fuhr ein Schneepflug und häufelte den Schnee am Straßenrand zu schon recht ansehnlichen Häufchen auf. Ich fuhr ganz langsam, aber wohl doch nicht vorsichtig genug: An einer Kreuzung verlor ich plötzlich die Kontrolle über den Wagen und ab ging es Richtung rechter Straßenrand. Kurz darauf fand ich mich in meinem Wagen mitten im Schnee wieder. Was für ein erfreulicher Tagesbeginn!

Auch wenn mir nichts weiter passiert war, erschienen, während ich auf Hilfe wartete, vor meinem inneren Auge die Schlagzeilen, mit denen in der morgigen Zeitung über das heutige Verkehrschaos berichtet würde: (...)

(175 words; Translation: MN)

AN ORDINARY WINTER'S TALE

Upon awakening one morning I noticed to my dismay that the first snows of autumn had falling during the night, and I didn't even have my winter tyres on the car yet, or even ready waiting in the garage. But the morning bus to town had already left, so I had to drive. I thought I would be able to get some proper tyres put on at the service station.

On the way the snow flurries turned to sleet, and **conditions** rapidly worsened. A snowplow ahead of me had left a low **snowbank** on the side of the road. I was driving slowly, but apparently not with sufficient care: at an inter-section the car simply began to slide toward the right shoulder, and I found myself and my car stuck in a **snowdrift.** A great way to start the day.

Even though I didn't hurt myself, I knew, while waiting for help to arrive, what kind of headlines about the morning's traffic would appear in the papers tomorrow. (. . .)(172 words; Translation: SC)

Appendix 2: Front page of a local Volkswagen customer leaflet. 2

A draft translation of the beginning:

"YOU CAN'T CHOOSE THE WEATHER CONDITIONS …WINTER IS COMING! Dark nights and rainy weather weaken visibility and headlights seem powerless. The first frosty morning can cause problems; door locks are frozen, the accumulator is empty, dampers don't work, the hand break is stuck etc. IS YOUR CAR READY FOR THE WINTER TO COME? […]"

References

Chesterman, Andrew (1993a). From 'is' to 'ought': laws, norms and strategies in translation studies. *Target, 5*(1), 1–20.

Chesterman, Andrew (1993b). Translation as theory. *Kääntäjä, 3*, 1.

Eskola, Sari (this volume). Untypical Frequencies in Translated Language: a Corpus-Based Study on a Literary Corpus of Translated and Non-Translated Finnish. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Fillmore, Charles J. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Linguistic Structures Processing* (pp. 55–88). Amsterdam: N. Holland.

Haarala, Risto et al. (1990–1994). *Suomen kielen perussanakirja* [Dictionary of Contemporary Finnish]. Vol. 1–3. Helsinki: Kotimaisten kielten tutkimuskeskus.

Holz-Mänttäri, Justa (1989). Interferenz als naturbedingtes Rezeptionsdefizit: ein Beitrag aus translatologischer Sicht. In H. Schmidt (Ed.), *Interferenz in der Translation* (pp. 129–134). Leipzig: Enzyklopädie.

Hönig, Hans G. (1988). Übersetzen lernt man nicht durch Übersetzen: Ein Plädoyer für eine Propädeutik des Übersetzens. *Fremdsprachen lehren und lernen, 17*, 154–167.

Hönig, Hans G. (1995). *Konstruktives Übersetzen*. Tübingen: Stauffenburg.

Hurme, Raija, Malin, Riitta-Leena, & Syväoja, Olli (1984). *Uusi suomi-englanti suursanakirja/Finnish-English General Dictionary.* Porvoo/Helsinki/Juva: WSOY.

Jääskeläinen, Riitta (1999). *Tapping the Process: An Explorative Study of the Cognitive and Affecitve Factors Involved in Translating* [University of Joensuu Publications in the Humanities 22]. Joensuu: University of Joensuu.

Jääskeläinen, Riitta, & Tirkkonen-Condit, Sonja (1991). Automatised processes in professional vs. non-professional translation: a think-aloud protocol study. In S. Tirkkonen-Condit (Ed.), *Empirical Research on Translation and Intercultural Studies. Selected Papers of the TRANSIF Seminar, Savonlinna 1988* (pp. 89–109). Tübingen: Gunter Narr.

Katara, Pekka, & Schellbach-Kopra, Ingrid (1997). *Suomi-saksa – suursanakirja/ Großwörterbuch Finnisch-Deutsch*. Porvoo, Helsinki, Juva: WSOY.

Kujamäki, Pekka (2003). Kaappaa käännösprosessi! Näytönkaappausohjelmistot aineiston-keruumenetelmänä ja kääntämisen opetuksen ja tutkimuksen työkaluna [Capture the screen! Screen recording applications as a method of data collection and as a pedagogical tool in translation class]. Paper presented at the KÄTU 2003 Translation Studies Symposium, April 4–5, 2003, Turku.

Kussmaul, Paul (2000a). A cognitive framework for looking at creative mental processes. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects* (pp. 57–71). Manchester, UK / Northampton, MA: St. Jerome.

Kussmaul, Paul (2000b). Types of creative translating. In A. Chesterman, N. Gallardo San Salvador. & Y. Gambier (Eds.), *Translation in Context: Selected Contributions fromt the EST Congress, Granada 1998* (pp. 117–126). Amsterdam/Philadelphia: John Benjamins.

Lörscher, Wolfgang (1987). Zur Genese und Entwicklung von Übersetzungskompetenz. *TextconText, 2*, 193–215.

Mauranen, Anna (2001). Strange strings in translated language: a study on corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects* (pp. 119–141). Manchester, UK / Northampton, MA: St. Jerome.

Mauranen, Anna (this volume). Corpora, Universals and Interference. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

Olohan, Maeve, & Baker, Mona (2000). Reporting *that* in translated English: evidence for subconscious processes of explicitation? *Across Languages and Cultures, 1*(2), 141–158.

Schmidt, Heide (1989). *Interferenz in der Translation*. Leipzig: Enzyklopädie.

Snell-Hornby, Mary (1983). *Metaphorical Thought and Translation: Taking a stand on P. Newmark*. Trier: L.A.U.T.

Tirkkonen-Condit, Sonja (1987). A pilot study into an aspect of the translation process. *Neuphilologische Mitteilungen, 88*, 221–229.

Tirkkonen-Condit, Sonja (2000). In search of translation universals: non-equivalence or 'unique' items in a corpus test. Paper presented at the UMIST/UCL Research Models in Translation Studies Conference, 28–30 April 2000, Manchester.

Tirkkonen-Condit, Sonja (2002). Process research: state of the art and where to go next. *Across Languages and Cultures, 3*(1), 5–19.

Tirkkonen-Condit, Sonja (this volume). Unique Items – Over- or Under-Represented in Translated Language? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* Amsterdam/Philadelphia: John Benjamins.

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

Toury, Gideon (1991). Experimentation in translation studies: achievements, prospects and some pitfalls. In S. Tirkkonen-Condit (Ed.), *Empirical Research in Translation and Intercultural Studies*: *Selected papers of the TRANSIF Seminar, Savonlinna 1998* (pp. 45–66). Tübingen: Gunter Narr.

Toury, Gideon (1995). *Descriptive Translation Studies and Beyond.* Amsterdam/Philadelphia: Benjamins.

Vannerem, Mia, & Snell-Hornby, Mary (1986). Die Szene hinter dem Text: 'scenes-and- frames semantics'in der Übersetzung. In M. Snell-Hornby (Ed.), *Übersetzungs-wissenschaft: Eine Neuorientierung* (pp. 184-205). Tübingen: Francke.

The fate of "The Families of Medellín"

Tampering with a potential translation universal in the translation class

Riitta Jääskeläinen

University of Joensuu

Avoiding repetition is one of the assumed translation universals, which professional translators (as good writers) tend to engage in almost automatically. However, sometimes repetition is used deliberately as a stylistic device. This article reports on a small-scale research project in progress which aims at finding out if and how students of translation can be made aware of the function of deliberate repetition in texts. The research material consists of student translations of the same source text from English into Finnish. The translation brief has been formulated such that the ST style ought to be preserved in the translation. Some groups of students have been asked to translate "blind", while others have been given instructions about style analysis and stylistic devices. A comparison of the students' translations indicates that students tend to avoid repetition, unless they have been sensitised to its importance as a feature of ST style.

1. Introduction

The present article is an interim report on a small-scale research project in progress, the aim of which is to find out whether the "avoidance of repetition" universal can be seen at work among translation students (i.e. novices); if yes, whether translation students could be weaned from automatic avoidance of repetition when necessary. The research project started from my own informal observations in translation class: students were presented with a source text which utilises repetition as a stylistic device and they were asked to translate the text for a purpose which called for a relatively "faithful" translation (the source text and the translation brief will be described in more detail below). To my surprise, repetition had been cleaned away from several translations;

as a result, some of the translations were clearly summarised versions. These observations made me look for an explanation for the students' behaviour and to think about finding a remedy.

One potential explanation for the students' behaviour is offered by one of the assumed universals of translation: the avoidance of repetitions which occur in the source text. Gideon Toury (1991:188) argues that the avoidance of repetitions is "one of the most persistent, unbending norms in translation in all languages studied so far". According to Toury (ibid.) avoiding repetition takes place "irrespective of the many functions repetitions may have in particular source texts," which is supported by my classroom observations. Toury's argument is also supported by research evidence from professional translation (e.g. Blum-Kulka and Levenston 1983, quoted in Laviosa-Braithwaite 1998:289; Toury 1991). It has been suggested that the (apparently universal) tendency to avoid repetition results from the assumed linguistic norms and rules of good writing, which the translators tend to follow as good professional text-producers.

Partly to test whether "avoidance of repetitions" is indeed at work among the novices and partly to try out a remedy (sensitising students to the stylistic functions of repetition), I decided to carry out a small research project in my translation courses.[1] In short, the idea was to ask one group of students to translate the text "blind," while another group would be given instructions on style analysis, and to find out whether any systematic differences could be identified between the two groups of translations.

In what follows, I will first introduce the research design, the source text, and the translation brief as well as the instructions on style analysis which were given to the students. Then I will discuss examples from my material. At this point of the research project I have looked at isolated (and random) examples of repetition in the ST and their translations to determine whether a more detailed analysis of the functions and translations of repetition in the whole text would make any sense. That is, the following observations do not relate to the whole text, but apply only to a few randomly selected examples.

2. Research design

The research material has been collected as part of the students' first course in translation from English into Finnish; the students who take the course are first-year students with English (translation and interpreting) as their major subject, which makes them clearly novices in translation. With the exception of

the translations collected in 2000, the Medellín translation has been one of the assessed translations in the translation course to ensure that the students take the task seriously. To ensure fairness, a particular group of students has always been given the same treatment as far the style instructions are concerned. The translations produced with instructions (N = 37) have been collected in 1996 and 2000, and the translations produced without instructions (N = 45) have been collected in 1996, 1999, and 2001. At present the material thus comprises a total of 82 translations.

The style analysis sheet was prepared by Kari Honkanen when he collected the with-instructions material in 1996. The instruction sheet has been compiled from different Finnish sources, and it contains passages illustrating the use of cohesive devices (including repetition and contrast) and describing different classifications of text-types. The instruction sheet may not be ideal, and not necessarily what I would use now, but the same instruction sheet must be used to keep the translating situations as similar as possible. (Obviously different teachers create different learning environments, which is an unfortunate confound at this stage of the project; on the other hand, different teachers help to level out the "teacher effect" on the results.) Nevertheless, the instruction sheet contains the relevant information without unduly underlining the features at the focus of my research interests. That is, the instruction sheet allows sufficient room for the students' creative thinking and problem-solving. The style analysis has been done in class before the deadline for the students' translations to ensure that the students have really paid attention to the instructions, but they have not been explicitly told what to do when they translate the text.

3. Source text analysis

The author of the source text, "The Families of Medellín", is Oscar Calle who was born in Medellín, Colombia, but was living in the US at the time the text was published in Newsweek on 14 March, 1988. The text is argumentative: the author wishes to make a point. The author wishes to personalise drug-related crime and drug-related deaths; while it is tempting to become immune to reports of violence abroad, the author wants to remind the readers that the distant victims of the drug trade are in fact somebody's loved ones, members of somebody's family. This can be illustrated by examples (1) and (2) from the source text. Example (1) is the second paragraph of the ST (the beginning of the first paragraph is presented in example (4) below) which comments on the assassination of a 28-year old man in Medellín.

(1)   The news of his assassination was hardly noted in the newspapers in his hometown of Medellín, Colombia; they are so used to it – 16 of these killings take place every single week in that city. It was not mentioned in the newspapers of other Colombian cities; this news is no news anymore in a country where 11,000 of these murders take place every single year. The international wire services didn't carry this event to their foreign affiliates; how many thousands of killings take place every single day in the world?

Three paragraphs later, in example (2), we return to the young murder victim who, as it turns out, was the son of one of the author's close friends. This fact is brought up after the author has established his own close personal relationship with Medellín.

(2)   I was born in Medellín. Many years ago I fell in love there. My first two children were born there. My father is buried there. I remember Christmas, birthdays, baptisms and funerals. And serenades at midnight. Medellín is also home for the so-called Medellín cartel, possibly the most powerful group of narcotraffickers in the world. It was the home, too, of the 28-year-old boy. His family is like my family. The father, Luis Fernando, is my friend who, not so long ago, drank aguardiente with me on nights of serenade. Last December he lost his son.

To make his point, the author uses an interesting variety of stylistic devices: he combines autobiographical narration with expository prose. He adds local colour by using Spanish loan words, such as *narcotraficantes*, *aguardiente*, and *corrida*. Repetition of various kinds is a central device, often coupled with contrasts. Repetition is also the means by which the author carries across the main image or metaphor in the text, i.e. *family* and *home*. The text is built on contrasts between good vs. bad families, the home of good vs. bad things/people (see example (2) above), families then and now, which is illustrated in the following excerpts from the ST.

The semantic network dealing with *family/home* is triggered by the headline "The *Families* of Medellín", and further enhanced in the caption "The city where I was born and where my father is buried is also the *hometown* of the cocaine cartel." The caption also expands the families mentioned in the headline into the good (=the author's) vs. bad (=drug dealing) families (i.e. contrasts the families). The caption also contains a couple of extensions of the family/home image, also contrasted, i.e. the author and his *birthplace*, the author's *father* and his *burial place*. (These references to family affairs are repeated later, as shown by example (2) above.)

The home/family idea is also given metaphorical extensions, as shown in example (3) below. The author is reminiscing about the Medellín he remembers from his childhood, and the passage contains the following paragraph.

(3) "The City of the Eternal Spring" and "The Beautiful Village" are two of the names that have been given to my hometown. When I was growing up that's exactly what it was, a city of beauty and charm located in a valley where the color green **must have been born**, and where rainbows **made their home**.

The text utilises lexical and structural repetition. For example, the beginning of the first paragraph uses both lexical repetition ('28') and anaphora ('One by one, ...'), as shown in example (4).

(4) **Twenty-eight** holes, **28** bullets, **28** years old. *One by one*, the **28** shots were fired with wrath from a short distance. *One by one*, they pierced the skin, ripped the flesh, tore the muscles, blew the vital organs away, and then with savage fury they exploded on their way out of the lifeless body, carrying with them a young man's dreams and tomorrows.

In sum, the author has utilised several stylistic means, including repetition, to carry across his highly personal message about the tragedy created by the drug trade in Colombia. As a result, the text offers interesting material for classroom experiments in translator training.

4. Student translations

In this section I will discuss randomly chosen examples from the student translations. The translation brief given to the students was formulated such that a relatively "faithful" translation was required; the brief was to translate the text to appear as a column in the Finnish quality weekly *Suomen Kuvalehti*. It would of course be possible to give the source text such a function in the target culture that the prominent features of style could be ignored in the translation, but here the purpose was initially to give the students an exercise in translation where preservation of style is essential.

My first example deals with the translation of the headline, "The Families of Medellín." As was mentioned earlier, the headline together with the caption act as triggers for the entire network of expressions related to *family/home*. As a result, it is important to retain the part of the headline ("families") which serves this function. Table 1 shows the results in this respect; the students'

Table 1. Back-translations of the headline

without instructions (N = 45)		with instructions (N = 37)	
The families of Medellín	24	The families of Medellín	27
Families of Medellín [PART]	1	Families of Medellín [PART]	1
Medellín, my hometown	2	The different families of Medellín	1
My hometown Medellín	1	The families of the city of Medellín	1
The children of Medellín	1	The families [EXT] of Medellín	2
		Family life in Medellín	1
		Family life [PART] in Medellín	1
The drug families of Medellín	1	Medellín is ruled by drug families [EXT]	1
Family life in the shadow of drugs	1		
total	**31**	**total**	**35**
Life in Medellín	3	Life in Medellín	1
The inhabitants of Medellín	1	Life in Medellín, the centre of the drug world	1
The city of Medellín	1		
Which way, Medellín?	1		
Medellín – the city of drugs	1		
The Medellín cartel	1		
Will drugs destroy Medellín ?	1		
Medellín – hell on earth	1		
The two faces of drug trade	1		
The price of drugs	1		
Rough game in Colombia	1		
Once upon a time in Colombia	1		
total	**14**	**total**	**2**

headlines have been back-translated from Finnish almost literally. Note that in Finnish there are two equivalents for the word "family"; one which refers to the immediate or nuclear family (*perhe*) and another one which refers to the extended family with uncles and aunts and cousins (*suku*). I have also kept apart the translations in which the word "families" is in the nominative case (*perheet*) or in the partitive case (*perheitä*). The translations which retain *family/home* are listed first, and below them the headlines where *family/home* has been lost.

Table 1 shows that the students translating with instructions have tended to retain the *family* in their headlines. In contrast, the students translating without instructions have produced a wealth of alternative formulations (which, admittedly, are often good descriptive headlines as such). Table 2 shows the percentages of the headlines which retain or lose the *family* connection.

Table 2. The distribution of the translations of the headline

without instructions (N = 45)	%	with instructions (N = 37)	%
metaphor retained (N = 31)	69	metaphor retained (N = 35)	95
metaphor lost (N = 14)	31	metaphor lost (N = 2)	5

An overwhelming majority (95%) of the students translating with instructions have kept *family* in the headline. Those translating without instructions have been more liberal in their choices: 31% of the headlines in this group do not retain *family*. As far as the repetition universal is concerned, the headline as a special case is of course slightly problematic. However, in terms of the effects of stylistic "sensitivity training" this example is rather encouraging.

The second example deals with the anaphoric sentences at the beginning of the first paragraph (see excerpt 2 above); the students' solutions are shown in Table 3.

Table 3. Translations of the anaphoric sentences in the first paragraph

ST: One by one, – One by one, –	without instructions (N = 45)	with instructions (N = 37)
Anaphora retained	27 = 60%	27 = 73%
Anaphora changed	18 = 40%	10 = 27%

The figures in Table 3 seem to point to a tendency to translate faithfully both with and without instructions, as the majority of students in both conditions have retained the anaphora. However, the instructions seem to have strengthened the tendency, which might indicate that the instructions have had the desired effect.

The third example of students' translation solutions deals with lexical repetition in the last sentence of the paragraph which is shown in example (5) (printed in bold).

(5)   The children today are not learning about beauty with Mistral. They don't need Joyce to teach them about girls; at 15 they know more than we ever dreamed. But they don't dream anymore. **They buy, they kill, they die.**

Table 4a shows the students' solutions to translating this sentence; Table 4b shows the percentages of translations in which the pronoun "they" has been repeated or changed.

Table 4b seems to offer support to both the repetition universal as well as the effect of remedial action. The majority (62%) of the students translating

Table 4a. Translation variants of pronoun repetition

Back-translations	without instructions (N = 45)	with instructions (N = 37)
They buy, they kill, they die.	13 = 29%	17 = 46%
They buy, they kill and they die.	4 = 9%	5 = 13.5%
They buy, kill, die.	8 = 18%	5 = 13.5%
They buy, kill and die.	9 = 20%	8 = 22%
Other	11 = 24%	2 = 5%

Table 4b. Percentages of translations retaining vs. changing pronoun repetition

ST: They buy, they kill, they die.	without instructions (N = 45)	with instructions (N = 37)
Pronoun repetition retained	17 = 38%	22 = 59%
Pronoun repetition changed	28 = 62%	15 = 41%

without instructions have tampered with pronoun repetition, while most of those translating with instructions have tended to retain it (59%). In this case sensitivity training seems to have reversed the students' tendency to avoid repetition.

My last example deals with the metaphoric extensions of *family/home*, which were mentioned in relation to example (3): "a valley where the color green *must have been born*, and where rainbows *made their home*." The figures in Table 5 show the student solutions which (1) retain both metaphors, (2) retain one of the metaphors, or (3) which have changed or omitted both metaphors. First I will give back-translated examples of each of the three types of solutions. (In Table 5 the number of translations produced with instructions is 35 instead of 37, as in two translations this page was missing due to a photocopying mishap; these two translations will be left out of the final analysis, but I have kept them in the material in the early stages of the project.)

a. **both metaphors retained:**
      where greenness must have be *born* and where rainbows had their *home*

b. **one metaphor retained:**
      where the green colour seems to have been *born* and where rainbows ended

c. **both metaphors changed or omitted:**
      in an evergreen valley where the sky was decorated by rainbows

Table 5. Retaining vs. changing the metaphoric instances of home/family

Home/family metaphor	without instructions (N = 45)	with instructions (N = 35)
both retained	22 = 49%	15 = 43%
one retained	16 = 36%	14 = 40%
both missing	7 = 15%	6 = 17%

With the repetition related to the metaphoric extensions the students' be-haviour seems to be more random than with the "normal" kinds of repetition in the ST, although the differences are not very great. The reason could be that we are dealing with a slightly different phenomenon here, and students do not identify the metaphorical extensions as such. On the other hand, this may also imply that the students' unit of translation is not large enough; students may operate successfully at the level of sentences or paragraphs; but they do not operate at the level of the whole text.

5.   Concluding remarks

The isolated examples of ST repetition and the students' reactions to it dis-cussed in this article give a somewhat incoherent picture. In some cases there is evidence of avoidance of repetition; in other cases the mechanism at work seems to be the principle of "faithful" translation (typical of first-year students who enter the university with a firmly rooted school-translation concept). In some cases stylistic sensitivity training seems to produce results, while some-times the students seem more immune. Obviously, as novices in translation the students have not yet internalised the unspoken "norms" of translation which professionals might share (cf. Blum-Kulka & Levenston 1983). Furthermore, as less experienced writers, they might not apply the "rules of good writing" as systematically as experienced writers, such as professional translators, do. In fact, there are a few intriguing exceptions in my material; these are students who have entered translator training after studying e.g. English philology for a couple of years. These students seem to be more inclined to avoid repetition than the genuine novices who have entered university right after school.

The findings may also stem from the fact that the examples deal with different kinds of phenomena. In the future it might make more sense to treat the repetition related to the metaphoric extensions separately from the "ordinary" kinds of lexical and structural repetition. There are also other

factors to be considered; as I mentioned earlier, the teaching environment appears to play a role, although I have not done a systematic comparison of the translations collected by myself and my colleagues. Furthermore, until now I have been working on the intuitive assumption that the same features of style operate in a similar fashion in both English and Finnish; this question needs to be addressed in the future. On the whole, however, I feel that the observations discussed here show that the classroom experiment in progress merits my attention also in the future.

Note

1. As teaching translation classes has not always belonged to my job description, I have also relied on the help of my colleagues to collect this material. I am very grateful to Kari Honkanen, Kati Martikainen and Tiina Puurtinen for their assistance.

References

Blum-Kulka, Shoshana, & Levenston, Edward A. (1983). Universals of lexical simplification. In C. Faerch & G. Kasper (Eds.), *Strategies in Interlanguage Communication* (pp. 119–139). London/New York: Longman.

Laviosa-Braithwaite, Sara (1998). Universals of translation. In M. Baker (Ed.), *Routledge Encyclopedia of Translation Studies* (pp. 288-291). London/New York: Routledge.

Toury, Gideon (1991). What are descriptive studies into translation likely to yield apart from isolated descriptions? In K. van Leuven-Zwart, & T. Naaijkens (Eds.), *Translation Studies: The State of the Art. Proceedings from the First James S. Holmes Symposium on Translation Studies* (pp. 179–192). Amsterdam: Rodopi.

Author index

Subject index

In the BENJAMINS TRANSLATION LIBRARY the following titles have been published thus far or are scheduled for publication:

1. SAGER, Juan C: *Language Engineering and Translation: Consequences of automation*, 1994.
2. SNELL-HORNBY, Mary, Franz PÖCHHACKER and Klaus KAINDL (eds): *Translation Studies: An interdiscipline. Selected papers from the Translation Congress, Vienna, 9–12 September 1992.* 1994.
3. LAMBERT, Sylvie and Barbara MOSER-MERCER (eds): *Bridging the Gap: Empirical research on simultaneous interpretation.* 1994.
4. TOURY, Gideon: *Descriptive Translation Studies — and beyond.* 1995.
5. DOLLERUP, Cay and Annette LINDEGAARD (eds): *Teaching Translation and Interpreting 2: Insights, aims, visions. Selected papers from the Second Language International Conference, Elsinore, 4–6 June 1993.* 1994.
6. EDWARDS, Alicia Betsy: *The Practice of Court Interpreting.* 1995.
7. BEAUGRANDE, Robert de, Abdulla SHUNNAQ and Mohamed Helmy HELIEL (eds): *Language Discourse and Translation in the West and Middle East. Selected and revised papers from the conference on Language and Translation, Irbid, Jordan 1992.* 1994.
8. GILE, Daniel: *Basic Concepts and Models for Interpreter and Translator Training.* 1995.
9. REY, Alain: *Essays on Terminology.* 1995.
10. KUSSMAUL, Paul: *Training the Translator.* 1995.
11. VINAY, Jean Paul and Jean DARBELNET: *Comparative Stylistics of French and English: A methodology for Translation.* 1995.
12. BERGENHOLTZ, Henning and Sven TARP: *Manual of Specialised Lexicography: The preparation of specialised dictionaries.* 1995.
13. DELISLE, Jean and Judith WOODSWORTH (eds): *Translators through History.* 1995.
14. MELBY, Alan with Terry WARNER: *The Possibility of Language. A discussion of the nature of language, with implications for human and machine translation.* 1995.
15. WILSS, Wolfram: *Knowledge and Skills in Translator Behavior.* 1996.
16. DOLLERUP, Cay and Vibeke APPEL: *Teaching Translation and Interpreting 3. New Horizons. Papers from the Third Language International Conference, Elsinore, Denmark 9–11 June 1995.* 1996.
17. POYATOS, Fernando (ed.): *Nonverbal Communication and Translation. New perspectives and challenges in literature, interpretation and the media.* 1997.
18. SOMERS, Harold (ed.): *Terminology, LSP and Translation. Studies in language engineering in honour of Juan C. Sager.* 1996.
19. CARR, Silvana E., Roda P. ROBERTS, Aideen DUFOUR and Dini STEYN (eds): *The Critical Link: Interpreters in the Community. Papers from the 1st international conference on interpreting in legal, health and social service settings, Geneva Park, Canada, 1–4 June 1995.* 1997.
20. SNELL-HORNBY, Mary, Zuzana JETTMAROVÁ and Klaus KAINDL (eds): *Translation as Intercultural Communication. Selected papers from the EST Congress – Prague 1995.* 1997.
21. BUSH, Peter and Kirsten MALMKJÆR (eds): *Rimbaud's Rainbow. Literary translation in higher education.* 1998.
22. CHESTERMAN, Andrew: *Memes of Translation. The spread of ideas in translation theory.* 1997.
23. GAMBIER, Yves, Daniel GILE and Christopher TAYLOR (eds): *Conference Interpreting: Current Trends in Research. Proceedings of the International Conference on Interpreting: What do we know and how?* 1997.

24. ORERO, Pilar and Juan C. SAGER (eds): *Translators on Translation. Giovanni Pontiero.* 1997.
25. POLLARD, David E. (ed.): *Translation and Creation. Readings of Western Literature in Early modern China, 1840–1918.* 1998.
26. TROSBORG, Anna (ed.): *Text Typology and Translation.* 1997.
27. BEYLARD-OZEROFF, Ann, Jana KRÁLOVÁ and Barbara MOSER-MERCER (eds): *Translator Strategies and Creativity. Selected Papers from the 9th International Conference on Translation and Interpreting, Prague, September 1995. In honor of Jirí Levi and Anton Popovic.* 1998.
28. SETTON, Robin: *Simultaneous Interpretation. A cognitive-pragmatic analysis.* 1999.
29. WILSS, Wolfram: *Translation and Interpreting in the 20th Century. Focus on German.* 1999.
30. DOLLERUP, Cay: *Tales and Translation. The Grimm Tales from Pan-Germanic narratives to shared international fairytales.* 1999.
31. ROBERTS, Roda P., Silvana E. CARR, Diana ABRAHAM and Aideen DUFOUR (eds): *The Critical Link 2: Interpreters in the Community. Papers from the Second International Conference on Interpreting in legal, health and social service settings, Vancouver, BC, Canada, 19–23 May 1998.* 2000.
32. BEEBY, Allison, Doris ENSINGER and Marisa PRESAS (eds): *Investigating Translation. Selected papers from the 4th International Congress on Translation, Barcelona, 1998.* 2000.
33. GILE, Daniel, Helle V. DAM, Friedel DUBSLAFF, Bodil MARTINSEN and Anne SCHJOLDAGER (eds): *Getting Started in Interpreting Research.* 2001.
34. GAMBIER, Yves and Henrik GOTTLIEB (eds): *(Multi) Media Translation. Concepts, practices, and research.* 2001.
35. SOMERS, Harold (ed.): *Computers and Translation. A translator's guide.* 2003.
36. SCHMID, Monika S.: *Translating the Elusive. Marked word order and subjectivity in English-German translation.* 1999.
37. TIRKKONEN-CONDIT, Sonja and Riitta JÄÄSKELÄINEN (eds): *Tapping and Mapping the Processes of Translation and Interpreting. Outlooks on empirical research.* 2000.
38. SCHÄFFNER, Christina and Beverly ADAB (eds): *Developing Translation Competence.* 2000.
39. CHESTERMAN, Andrew, Natividad GALLARDO SAN SALVADOR and Yves GAMBIER (eds): *Translation in Context. Selected papers from the EST Congress, Granada 1998.* 2000.
40. ENGLUND DIMITROVA, Birgitta and Kenneth HYLTENSTAM (eds): *Language Processing and Simultaneous Interpreting. Interdisciplinary perspectives.* 2000.
41. NIDA, Eugene A.: *Contexts in Translating.* 2002.
42. HUNG, Eva (ed.): *Teaching Translation and Interpreting 4. Building bridges.* 2002
43. GARZONE, Giuliana and Maurizio VIEZZI (eds): *Interpreting in the 21st Century. Challenges and opportunities.* 2002.
44. SINGERMAN, Robert: *Jewish Translation History. A bibliography of bibliographies and studies.* 2002.
45. ALVES, Fabio (ed.): *Triangulating Translation. Perspectives in process oriented research.* 2003.
46. BRUNETTE, Louise, Georges BASTIN, Isabelle HEMLIN and Heather CLARKE (eds): *The Critical Link 3. Interpreters in the Community. Selected papers from the Third International Conference on Interpreting in Legal, Health and Social Service Settings, Montréal, Québec, Canada 22–26 May 2001.* 2003.

47. SAWYER, David B.: *Fundamental Aspects of Interpreter Education. Curriculum and Assessment.* n.y.p.
48. MAURANEN, Anna and Pekka KUJAMÄKI (eds): *Translation Universals. Do they exist?* 2004.
49. PYM, Anthony: *The Moving Text. Localization, translation, and distribution.* n.y.p.
50. HANSEN, Gyde, Kirsten MALMKJÆR and Daniel GILE (eds.): *Claims, Changes and Challenges in Translation Studies. Selected contributions from the EST Congress, Copenhagen 2001.* n.y.p.
51. CHAN, Leo Tak-hung: *Twentieth Century Chinese Translation Theory. Modes, issues and debates.* 2004.