

Lucja Biel*

Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential

Abstract

The introduction of electronic corpora to linguistics has been compared to the introduction of telescopes in astronomy (Stubbs 2004: 107). Indeed, the use of computers in language studies was inevitable. Corpus-based and corpus-driven studies of language have become a mainstream methodology used in many branches of linguistics. Since the 1990s they have also been applied in translation studies, yet relatively seldom in research on specialised translation, in particular legal translation. The aim of this paper is to demonstrate the potential of corpus-based studies as a methodology for researching legal translation and as a tool in translator training.

1. Corpus-based studies of language as a methodology

The term **corpus linguistics** is relatively new as it dates back to the 1980s. Yet this methodology was known much earlier in a paper form. Svartvik (2007) and Stubbs (2004) describe its ‘Stone Age’ and list a few examples of ‘language corpora BC’ (before computers). The most notable one is a corpus of 5 million citation slips compiled by volunteers in the second half of the 19th c. and at the beginning of the 20th c. for the *Oxford English Dictionary* published in 1928 (see Stubbs 2004: 110). Other famous linguists who used ‘shoebox corpora’ included: Jespersen, Boas, Sapir, Fries, Bloomfield, and Pike (McEnery et al. 2006: 3). The fifties brought about a move away from empirical methods in favour of the rationalism paradigm, following Chomsky’s interest in linguistic competence and related scepticism of corpora due to their ‘skewedness’ (McEnery et al. 2006: 3). Despite the unsupportive attitude of linguistic circles it was in the 1960s when the first modern machine-readable corpus, the Brown University Standard Corpus of Present-Day American English, was built by Henry Kučera and W. Nelson Francis. It had only a million words and at that time its processing required the application of all available computer resources (Svartvik 2007: 20). Following developments in computer science, corpus-based studies of language re-emerged on a greater scale in the late 1980s, spreading into all possible areas and branches of linguistics and related disciplines. Its popularity may be confirmed by the sheer number of books and articles published on the topic. With time corpora have markedly increased in size: the Oxford English Corpus has 2 billion words, the Corpus of Contemporary American English has 400+ million words and the British National Corpus has 100+ million words.

What is a corpus? It is often defined as a machine-readable representative collection of naturally occurring language assembled for the purpose of linguistic analysis. Its analysis lays the foundations for corpus linguistics. Corpus linguistics is not, as Gries argues, a homogeneous methodology: it is used with a varying level of granularity and varying reliance on quantitative and qualitative methods, with its shared features being as follows: machine-readable naturally-occurring language, balanced and representative corpus design, systematic and exhaustive analysis:

- the analysis is based on a corpus or corpora of naturally-occurring language which are machine-readable so that the retrieval of the search patterns is computerized;

* Dr. Lucja Biel
Department of Translation Studies
Institute of English
University of Gdansk
Gdansk
Poland
anglb@univ.gda.pl

- the corpus is intended or taken to be balanced and/or representative of the modality/register/variety the study is aimed at;
- the analysis is, or at least attempts to be, systematic and exhaustive, meaning that the corpus does not simply serve as a database of examples from which some can be chosen *ad libitum* and others neglected but that the whole (sample of the) corpus is taken into consideration [...]. (Gries 2006: 4)

The last feature is related to one of Stubbs' principles of corpus-based studies — namely, “The observer must not influence what is observed”. His second principle, “Repeated events are significant”, is crucial for its approach to meaning. Corpora show what is “central and typical, normal and expected”; they emphasise that language use is highly patterned and that such patterns are not accidental but cognitively motivated (Stubbs 2004: 111). All combined, it promotes the theory of “meaning as use” as developed by Wittgenstein, Austin and Firth (Stubbs 2004: 110). Another important influence of corpora is that they take linguistics “beyond the single word as the basic semantic unit” (Teubert 2002: 212), posing an important theoretical question as to the minimum linguistic unit. Taylor, a cognitive linguist, notes that it might be more appropriate to use the term ‘mental phrasicon’ rather than ‘mental lexicon’ since the linguistic description of words must include constructions they occur in (2006: 575).

On the whole, the main advantages of corpus-based studies of language are as follows: reduced speculation and subjectivity; authenticity of data; and the potential to verify research hypotheses systematically and based on more extensive linguistic material. Its disadvantages include problems with representativeness and balance: any claims and generalisations we make about language are representative of the language sample we research, not of the entire language.

As a method, corpus linguistics gives priority to observation over intuition (Stubbs 2004: 107-8). It is classified as an **empirical** approach to the description of language use which analyses authentic data and is inductive/data-driven in that it formulates theoretical statements from observations of actual use (Tognini-Bonelli 2001: 2). It is mainly a quantitative method but it also integrates qualitiveness to hypothesise about data provided by the corpus and to form generalisations about language use (quantitative-driven qualitiveness); in this sense it is complementary with an intuition-based approach (McEnery et al. 2006: 7), or with what Fillmore calls armchair linguistics (Stubbs 2004: 108).

There is some controversy whether corpus linguistics is a methodology or a theory. The prevalent view is that it is not a theory or an independent branch of linguistics: as emphasised by Johansson, “it is not defined by the object of study ... the object of corpus linguistics is *not* the study of corpora. It is rather the study of language through corpora” (qtd. in Kenny 2001: 23). It is mainly regarded as a methodology which has developed its own systematic methods and principles for the application of corpora to studies of language use; hence, it is a methodology with a “theoretical status” used in many areas and theories of linguistics (McEnery et al. 2006: 7-8). For example, it is applied to the description of various areas of language (descriptive linguistics): semantics (collocations, colligations), syntax (corpus-based grammars), pragmatics (register variation, genre analysis, stylistics). Findings of corpus-based studies have been applied in various branches of linguistics and theoretical frameworks: lexicography (corpus-based dictionaries), sociolinguistics, applied linguistics (language learning), diachronic studies, discourse analysis, cognitive linguistics, as well as what we are mostly interested in: forensic linguistics, contrastive and comparable linguistics, and translation studies.

2. Types of corpora

Legal translation may benefit from corpus linguistics in a number of ways: it may be applied for theoretical and practical purposes. The way corpus linguistics is used is directly related to the type of corpora. From the point of view of translation studies, there are four major types of corpora: 1. monolingual corpus, 2a. monolingual comparable corpora, 2b. bilingual comparable corpora, 3. parallel corpus. Table 1 presents a summary of differences between the corpus types.

Type of analysis	Intralingual		Interlingual (cross-linguistic)	
Number of languages	Monolingual (1 language)		Bilingual/multilingual (2+ languages)	
Corpus design	(1) MONOLINGUAL 1 corpus	(2a) COMPARABLE 2+ corpora	(2b) COMPARABLE 2+ corpora	(3) PARALLEL 2+ corpora
Type	typical linguistic corpus	translation-driven corpus	translation-driven corpus	translation corpus
Number of languages	1 language	1 language	2 or more languages	2 or more languages
Corpus content	non-translated language A	translated versus non-translated language A	non-translated language A and B	non-translated language A aligned with translation in B
What may be examined	legal language against other genres	translated language against non-translated one	differences and similarities between languages	translation process

Table 1. Classification of corpora for translation purposes

1. **Monolingual corpus:** it is the most typical corpus used by linguists. It contains non-translated texts created only in one language. It involves intralingual analysis, within a single language, for example for descriptive purposes, but also to compare legal language against everyday language or other genres if a reference corpus is used. This type of corpus is mainly used within forensic linguistics or legilinguistics, but also in monolingual lexicography and in foreign language teaching to prepare study materials, as is the case with the Cambridge Corpus of Legal English, a 20-million-word collection of legal books and newspaper articles compiled by Cambridge University Press.
2. **Comparable corpora:** It is a set of at least two monolingual corpora which may involve one language (a) or at least two languages (b). Zanettin refers to them as “translation-driven corpora” since their design is motivated by translation research or training yet they do not contain source texts (STs) and corresponding target texts (TTs) (2000: 106).
 - a. **Monolingual comparable corpora:** they contain a corpus of translations and a corpus of texts created spontaneously in the same language (non-translated language). The main object of analysis is how the translated language differs from the non-translated language (to be discussed later as the ‘textual fit’). An example of such corpora is the Translational English Corpus at the University of Manchester. This type of corpora is used in translation studies.
 - b. **Bilingual or multilingual comparable corpora:** they do not contain translated language but spontaneously created texts in two different languages. It is a set of two monolingual corpora designed according to a similar criterion and is used for cross-linguistic analysis. In addition to translation studies, this type of corpora is typically associated with contrastive and comparable linguistics. An example of comparable corpora is the BOnonia Legal Corpus, BoLC, at the University of Bologna, with the Italian legal subcorpus of 33.5 m words and the English legal corpus of 21 m words¹.

¹ Data from the BoLC website: http://corpora.dslo.unibo.it/bolc_eng.html.

3. **Parallel corpus**² is a translation corpus in the strictest sense. It is bilingual or multilingual and may be bi-directional. It contains STs aligned with their translations. Alignment makes parallel corpora more time-consuming to build and, as a result, they are rather seldom found. Examples³ include: the MultiJur Multilingual Corpus of Legal Texts at the University of Helsinki, legal sections of the CLUVI Parallel Corpus at the University of Vigo (Galician-Spanish, Basque-Spanish) and the GENTT Corpus of Textual Genres for Translation at the Jaume I University. This type of corpus is mainly used for research into the translation process and in applied translation studies: to prepare dictionaries, extract terms for terminological databases, train information extraction software, and train translators.

One of the major limitations concerning the application of legal corpora is their availability, in particular for less researched languages. This claim is especially valid for parallel corpora⁴, which, as already noted, require time-consuming alignment of STs and TTs. Furthermore, accessibility to the existing corpora is limited due to copyright restrictions. Another issue which significantly impacts the accessibility, size and composition of legal corpora is confidentiality of legal documents, in particular private legal and litigation documents. Such texts tend to be excluded or are included in relatively small samples; hence, legal corpora — like most specialised corpora — are rather small. It is inevitable that legislation is over-represented while other legal genres are under-represented. In this sense corpora strengthen the ‘legicentrism’ that legal scholarship has been accused of (Kasirer qtd. in Harvey 2002: 178).

As regards the size of specialised corpora and involvement of rare language pairs, a major breakthrough is the Acquis Communautaire, the largest multilingual corpus of EU legislation and ECJ rulings, made publicly accessible by the European Commission. Two variants are available: JRC-Acquis and DGT’s Translation Memory of the Acquis. The current version 3.0 of JRC-Acquis⁵, an aligned multilingual parallel corpus, is available in 22 languages (excluding Irish), with the overall size of 1 bn words. It contains more than 23,000 documents per language with 55 m words for English and 50 m words for Polish. It was first released in 2006 and version 3.0 became available in April 2007. The DGT’s Translation Memory of the Acquis⁶ contains parallel texts for 22 EU languages. It allows users to prepare bilingual aligned corpora for one of 231 language pairs. These resources of an unprecedented scale may open up new vistas for translation scholars.

3. Trajectories of research in corpus-based studies of legal language

This section will survey major trajectories of corpus-based research on legal language contributed mainly from such fields as forensic linguistics, legilinguistics, contrastive and comparable linguistics. This type of research is based on monolingual and comparable corpora which contain non-translated language since its objective is to study legal language *per se*. Corpus-based studies of non-translated legal language may be grouped into the following major trajectories of research:

Trajectory 1: **External variation**: how does legal language differ from general language and other languages for special purposes?

2 The terminology for corpus types has been adopted after Zanettin (2000). It should be noted that the terms parallel corpus and comparable corpus are sometimes used interchangeably. Parallel corpora are also termed by some researchers as a *translation corpus*.

3 A non-exhaustive list of parallel corpora is available at: <http://tcc.itc.it/people/forner/multilingualcorpora.html>.

4 Some linguists argue that this claim is also true for monolingual corpora, which are much more frequent and easy to build. Blackwell in her article “Why forensic linguistics needs corpus linguistics” makes a plea for creating accessible specialised corpora for forensic linguistic purposes as those which are available right now tend to be too general (2009: 14).

5 <http://langtech.jrc.it/JRC-Acquis.html>

6 <http://langtech.jrc.it/DGT-TM.html>

Trajectory 2: **Internal variation**: how do legal genres differ from each other?

Trajectory 3: **Temporal variation**: how does the current legal language differ from a historic one?

Trajectory 4: **Cross-linguistic variation**: how does it differ across languages?

The first three trajectories are the main domain of research within forensic linguistics — a branch of linguistics which studies language used in the justice system and, more broadly, language and the law. In most cases forensic linguistic research is not translation-driven but serves other purposes. For example, corpus methodologies are used to attribute authorship of linguistic evidence in witness statements and ransom, blackmail or suicide notes, as well as to detect plagiarism (McEnery et al. 2006: 116). Apart from attempts to identify the “linguistic fingerprint” (Stubbs 2004: 124), forensic linguists are engaged in general studies of legal language.

Let us now demonstrate how monolingual corpora may be used. A good illustration of Trajectory 1 is a study of legal English against general English, conducted by Coulthard/Johnson (2007), who used the COMET corpus of legal contracts and the British National Corpus (BNC), respectively. Their findings show, *inter alia*, that legal language has a different distribution of not only lexical items but also, which may be more difficult to notice, of grammatical items. The latter include a distinctive distribution of *or*, *any*, *shall*, *be* and *by* which is markedly higher than in general language, *or* being explained by the wish to ensure inclusiveness while *by* is explained by passive constructions (2007: 40). As regards discrepancies in lexical distribution, while the 56 most frequent items identified in the BNC are grammatical, the COMET corpus includes as many as 15 lexical items in the 57 most frequent lexemes, e.g. *agreement*, *company*, *lessee*, *party*, *agent*, *notice*, *property*, etc. (2007: 44). As a result, contracts — and legal language in general — have been found to show a higher lexical density than general language (2007: 44).

Trajectory 2 (internal variation) spans a broad spectrum of studies, ranging from micro- to macroanalyses. The former dissect legal language into smaller units for descriptive purposes while the latter cover attempts to identify distinctive features of legal genres to show how they differ from each other. Since corpus methodology is ideally suited for researching phraseology, it is a frequent object of microanalysis. This may be illustrated by the present author’s ongoing project of the *Dictionary of Polish Legal Collocations for Translators*. It is based on my Polish Law Corpus of ca. 4 m words, which includes 211 codes and major legal acts related to contract, company, civil and criminal law. The objective is to describe nominal, verbal and adjectival collocations of legal terms to aid translators in achieving naturalness in target texts, collocations and colligations contributing significantly to the overall ‘feel’ of translation. Another example is Goźdz-Roszkowski’s analysis of lexical bundles in English contractual instruments with a view to discovering recurrent patterns, such as referential bundles, text organisers, and modalising bundles (2006). Macroanalysis helps identify legal genres. As suggested by Blackwell, corpora may also be applied to detect differences between the language of prosecution and defence lawyers or expert witnesses and eye-witnesses (2009: 14). The latter falls into a somewhat fuzzy area of the *legal language* category, that is expert-lay communication, which has been investigated by Heffer in his corpus-aided study of language used by legal professionals in communication with lay juries based on transcripts from ca. 100 criminal jury trials (2005).

Corpora are used not only in synchronic but also diachronic studies (Trajectory 3). A fascinating example of diachronic corpora is a digitised collection of the Proceedings of the Old Bailey, London’s Central Criminal Court (www.oldbaileyonline.org), which contains as many as 200,000 criminal trials from the period 1674–1913. Another such corpus is the Helsinki Corpus of English Texts: Diachronic and Dialectal, used for example by Gotti to research semantic and pragmatic values of the modals *shall* and *will* in Early Modern English statutes (2001).

Trajectory 4, cross-linguistic variation, is based on comparable corpora with components in at least two languages. “Seeing through corpora we can see through language,” argues Johans-

son (2007: 51), who emphasises that multilingual corpora move the researcher closer to the nature of an individual language: “the comparison sharpens the perception of both similarities and differences” (2007: 55). This claim may be substantiated by Carvalho’s study of binomial expressions, such as *terms and conditions*, *successors and assigns*, *costs and expenses*, *representations and warranties*, on her comparative corpus of English and Brazilian agreements of 140 texts each. Since these binomials tend to be translated literarily into Brazilian Portuguese rather than be treated as a unit, the corpus was designed to raise translators’ awareness of the difference (2007: 109).

The most comprehensive yet somewhat sceptical assessment of corpus applicability to studies of legal language is Bhatia, Langton and Lung’s paper entitled “Legal discourse: Opportunities and threats for corpus linguistics” (2004). The authors argue that legal language, in particular legislation, does not often need large corpora since, because of its conservatism and ‘formulaic form-function correlations’, it may be ‘equally efficient and reliable’ to conduct a manual analysis, for example on a single legislative act (2004: 207). Hence, “there is very little need for comprehensive or automatic linguistic frequency measures, as they are easily identifiable manually” (2004: 212) and “there is very little that corpus-based linguistic analysis of legal discourse will bring to light in this sense” (2004: 212). On the other hand, the authors agree on the usefulness of corpora in researching ‘intertextuality within and across a particular genre’, in particular to develop grammars of legal genres, which otherwise would be ‘tedious, inaccurate and incomplete’ (2004: 212); and they argue for the use of genre-based small corpora rather than a large corpus (2004: 215). In fact, what seems to be the most contested area is the possibility to research interdiscursivity, i.e. embedding of one genre into another, with the corpus methodology (2004: 222). It is claimed that in order to get a full picture of a genre, it is also necessary to study conventions that enable such embeddings and, for this purpose, to account for institutional, social and cognitive factors, which requires qualitative rather than quantitative analysis (2004: 224). The authors seem to take a narrow view of corpus linguistics, equating it with a quantitative approach: “one needs to appreciate that qualitative analysis begins where corpus linguistics ends” (2002: 224), even though corpus linguists acknowledge openly that it is necessary to combine quantitative methods with qualitative ones (quantitative-driven qualitiveness). And rather than excluding corpora, it is perhaps worth bearing in mind, after Johansson, that “it is a common experience among those who work with corpora that we often make new discoveries” (2007: 55).

4. Advances in corpus-based translation studies and their implications for legal translation

Legal translation research has traditionally focused on the system-bound nature of legal terms, related incongruity and limits of translatability, with relatively little interest in corpus-based methods. Applications of corpora were described in none of the three major textbooks on legal translation: Šarčević’s *New Approach to Legal Translation* (1997) and Alcaraz/Hughes’s *Legal Translation Explained* (2002), which given their dates of publication is not so surprising, but also not in the most recent one, Cao’s *Translating Law* (2007), except for a short passage on CAT tools. Since legal translation studies is an interdiscipline which is situated on the interface between translation studies, linguistics, terminology, comparative law, and cultural studies, it is impossible to ignore significant developments within translation studies stimulated by corpus-based research.

Corpus-based translation studies were developed in the mid 1990s and have continued to be intensely applied in the last decade. They mark a shift from the analysis of the ST-TT relation (i.e. equivalence, accuracy) to TTs as independent texts on their own, emphasizing the importance of translated texts in receiving cultures. This shift from the ST to the TT has been referred to by Pym as a ‘paradigm shift’ in translation studies (2004). It has contributed to the polysystem theory, descriptive translation studies and skopostheorie (Baker 1993). Although accuracy and faithfulness have been and will be a priority in legal translation, traces of TT reorientation may also be observed in legal translation research, with the emancipation of the legal translator (Šarčević 1997:

112) and a legal translation being perceived as an independent text which may “function on its own in the new situation without necessary recourse to the source text” (Engberg 2002: 382).

Focus on target texts drew attention to translated language as such. The early 1990s saw the accelerated development of corpus-based studies in linguistics, which, however, tended to exclude translations as a non-representative language and, in particular, exclude them from monolingual corpora (Baker 1996: 175). Recognising the advantages of corpus-based methodology, Baker pioneered its application in translation studies in her seminal papers (1993, 1996), where she proposed to analyse translations against non-translated texts and identify distinctive features of translated texts using corpus data. This proposal has been eagerly taken up by translation scholars, who with time “have come to rely on corpora to verify, refine or clarify theories that hitherto had had little or no empirical support and to achieve a higher degree of descriptive adequacy” (Granger 2003: 19).

The idea that translated language may be different from non-translated language is not new. The former has been referred to in the literature as: translationese, hybrid language, third code, third language, or translanguaging, the first term being most widespread. **Translationese** is the language of translation usually understood pejoratively as “translation-based deviations from target language conventions” (Doherty qtd. in Olohan 2004: 29) or as “translated language that appears to be influenced by the source language, usually in an inappropriate way or to an undue extent” (Olohan 2004: 90). It is an intermediary, hybrid language between the SL and the TL since a translated text “is essentially a third code which arises out of the bilateral consideration of the matrix and target codes: it is, in a sense, a sub-code of each of the codes involved” (Frawly 1984: 168). As noted by Shamma, an untypical distribution of lexical items may contribute to the translation-like impression and “leave[s] a vague impression of being culturally exotic” (qtd. in Baker 1993: 245). Obviously, the markedness of translationese may be higher in inexperienced or incompetent translators or in inverse translation into one’s non-native language where it is coupled with non-native competence.

The distinctive nature of translationese is caused not only by deviations from the norm due to SL interferences but also, more interestingly, by distinctive features of the translation process itself referred to by Baker as **translation universals**. She sees translation as “a mediated communicative event” (1993: 243), “shaped by its own goals, pressures and context of production” (1996: 176). In consequence, translations are believed to show unique recurrent features independent of language pairs, genres, cultures, etc. According to Baker, they include:

- **explicitation**: translators’ tendency to explicate what may be implicit in the ST,
- **simplification** and **disambiguation**: the tendency to simplify the message and/or language in TTs,
- **normalisation**/conservatism: the tendency to exaggerate typical features of the TL, and
- **levelling out**: the tendency of translations “to gravitate towards the centre of a continuum” as translations are less idiosyncratic and more similar to each other than original texts (1996: 180-185).

With time the list has been extended to include other features, such as under-representation of unique TL items or untypical collocations (Mauranen 2006: 95). Seeking generalisations about translation, Chesterman divides universals into S-universals and T-universals: the former concern the equivalence relation which holds between STs and TTs while the latter contribute to the textual fit between translations and naturally occurring non-translated language (2004: 6-7). This classification neatly reflects two crucial interrelated aspects of any specialised translation: accuracy and naturalness, respectively (Table 2).

S-universals THE EQUIVALENCE RELATION	T-universals THE RELATION OF TEXTUAL FIT
<ul style="list-style-type: none"> • lengthening of TTs compared to STs • Toury's law of interference • Toury's law of standardisation • dialect normalisation • reduction of complex narrative voices • explicitation • sanitization • retranslation hypothesis • reduction of repetition 	<ul style="list-style-type: none"> • simplification (lower lexical variety and lexical density, more high-frequency items) • conventionalisation, normalisation • untypical and less stable lexical patterning • under-representation of TL-specific items

Table 2. Chesterman's classification into S-universals and T-universals

The concept of translation universals may be controversial but it has elicited a heated response from translation scholars, pushing research in new directions and providing evidence both for and against their existence. The prevalent view is that they have the status of hypotheses. As emphasised by Chesterman, "Genuine universals are the subject of unrestricted hypotheses: these claims aim to be valid for all translations of all kinds, in all times and places, universally" (2004: 9), which makes them difficult, if not impossible, to prove. Major reservations concern the very term 'universal' and whether universals really arise from the translation process itself. House, who questions their existence, argues that they are language universals applicable to translation rather than autonomous translation universals (2008: 11). Her other counterarguments include evidence that features of translated texts may differ depending on the language pair, directionality of translation and genre; for example, a higher degree of explicitation was found in German translations of popular science texts than in economic texts (2008: 12). This leads House to argue that "the quest for translation universals is in essence futile, i.e. that there are no, and there can be no, translation universals" (2008: 11). Chesterman, who is sceptical as to the very possibility of proving the hypothesis, takes an opposite view on its usefulness: "What ultimately matters is perhaps not the universals, which we can never finally confirm anyway, but new knowledge of the patterns, and patterns of patterns, which help us to make sense of what we are looking at" (2004: 11). Therefore, the major contribution of research on translation universals is that it brings to light recurrent patterns likely to be found in translation. It is reasonable not to treat them as absolute laws but "general or law-like tendencies, or high probabilities of occurrence" (Mauranen 2006: 94).

Translation universals elicit a number of questions, still unanswered, concerning their potential impact on legal translation. Suppose the universals exist. Should we be concerned about the simplification or disambiguation tendency in light of the strategic ambiguities and purposeful flexibility of legal language? Is sanitation in conflict with the high accuracy requirement in legal translation? Given the levelling-out and untypical distribution, how will the large inflow of EU translations influence vernacular legal languages? And, more importantly, can our awareness of such tendencies improve the accuracy and naturalness of translations? Below is a discussion of relevant universals which require testing on legal translation.

4.1. Explicitation

Explicitation is defined by Baker as "an overall tendency to spell things out rather than leave them implicit in translation", which may be manifested at various levels: longer length of translations, overuse of explanatory vocabulary and conjunctions, rise in cohesion, etc. (1996: 180-181). According to Toury, explicitation appears in "all kinds of mediated events, including interaction in a foreign language" (qtd. in Baker 1993: 244). Blum-Kulka, who was first to formulate the explicitation hypothesis and observe the rising explicitness in translationese, notes that "the process of interpretation performed by the translator on the source text might lead to a TL text which is

more redundant than the SL text” (1986: 19). This tendency also seems to appear in legal translation but it may have an additional cause. Based on what is known about legal language, it may be hypothesised that explicitation will result from the need to bridge knowledge gaps between legal systems. One of the fundamental problems in legal translation is the incongruity of concepts between legal systems and the absence of universal knowledge structures, which are present in other types of specialised translation, e.g. in mathematics, biology, chemistry or technology (Biel 2008: 22). The degree of explicitation is connected with the conceptual distance between legal systems, the need for explicitation being higher when translation takes place between common-law and civil-law systems rather than between systems with similar legal traditions. Additionally, the degree of explicitation varies with the translator’s preference for SL-oriented (foreignising) or TL-oriented (domesticating) strategies. However, such preference is not always purely idiosyncratic and/or adjusted to the *skopos* of translation, but is also influenced by the culture-specific dominant school of thought amongst trainers, language-specific textbooks or, in some cases, by country-specific recommendations for legal translators. The influential Polish Sworn Translator Code, prepared in collaboration with the Polish Ministry of Justice, advises sworn translators that they have the right to assume that, as an expert, the recipient is aware of the incongruity of legal systems; hence, no additional explanations or definitions need to be provided (Kierzkowska et al. 2005: 87–88). In TL-oriented strategies, such as Šarčević’s receiver-oriented approach, translators are expected to “compensate for conceptual incongruity whenever possible” to ensure that the ST and the TT have the same legal effect (2000).

At a micro-level of textual procedures, translators apply a number of techniques to deal with incongruous terms, ranging from transcription (borrowing) and naturalisation, literal equivalents, descriptive equivalents to functional equivalents (see Biel 2009 for more detailed discussion). While functional equivalents are most TL-oriented, they are not always possible due to a high degree of incongruity or zero equivalence; in such a case, the translator may want to resort to a descriptive equivalent (also known as a gloss or a paraphrase). This technique attempts to account for the recipient’s knowledge gaps by explaining a concept which is absent in the target legal system (zero equivalence) or by explicating central aspects of meaning which are responsible for the incongruity. As regards the former, a descriptive equivalent is an alternative to a pure borrowing (transcription) with varying degrees of explicitation available. An example which suggests itself is *equity*: it may be retained as *equity*, preceded by a hyponym (*system prawa*), which explains that it is a system of law, or explicated fully as *system prawa słuszności* or *system oparty na zasadach słuszności*. The degree of explicitation is limited by practical considerations: terms function as shortcuts to knowledge structures and to be convenient in use they should be short (Biel 2009: 185). In the case of incongruous concepts, a functional equivalent may be modified to explain the crucial difference. Take for example *spółka jawna*, the most basic Polish partnership, whose recognised equivalent is *registered partnership*, which emphasises that, in contrast to the UK entity, the Polish partnership requires registration (this fact is implicit in Polish).

Explicitation is not limited to a few system-specific terms but also appears in the case of relatively congruous concepts. Many such examples may be found in the *DGT Polish Language Department In-house Glossary*⁷, which contains authoritative equivalents of EU-related terms. Compare for example *legal aid and advice* in the European Convention on the Exercise of Children’s Rights and its Polish equivalent of *prawo do przedstawiciela z urzędu i bezpłatnego doradztwa prawnego* (lit. the right to a court-appointed counsel and free-of-charge legal advice). Another striking example is *interline agreement*, which, upon consultation with an expert, is recommended by the DGT to be translated as *porozumienie o wzajemnym honorowaniu dokumentów przewozowych* (lit. understanding on mutual recognition of transport documents) rather than the previous literal – and quite comprehensible – equivalent of *umowy międzyliniowe*. While *negative price* has a French equivalent of *prix négatif* and a German one of *negativer Preis*, its Polish equi-

7 http://ec.europa.eu/translation/language_aids/freelance/polish_en.htm.

valents are less stable and vary in the degree of explicitation from literal equivalents of *cena negatywna* and *cena ujemna* to a metaphoric equivalent of *cena drapieżna* (lit. predatory price), and explicit ones: *cena poniżej wartości rynkowej lub księgowej* (lit. price below the market or book value) or *cena poniżej kosztów* (lit. price below costs), depending on the context and the local institution involved. Explicitation is even present in non-legal terms, e.g. in two equivalents of *neutral spirits* — *alkohol etylowy neutralny* (lit. neutral ethyl alcohol) and *alkohol etylowy pochodzenia roślinnego* (lit. ethyl alcohol of plant origin). It seems that explicitation is used not only to bridge knowledge gaps in the case of incongruous concepts, but also to enhance comprehension as explicit coding is easier to process than implicit information.

Yet this analysis reveals neither the scale of explicitation in actual translation practice nor factors that may reduce or increase explicitation. To answer these questions, parallel corpora need to be studied. One of the factors, as argued by House, is directionality (2008: 12). For example, translation from synthetic languages, such as Polish and German⁸, may involve some degree of implicitation, as in *data wystawienia faktury* (lit. date of issuing of invoice), which should be translated implicitly as *date of invoice*. This direction may enable two options: a literal equivalent with the same or higher degree of explicitation or an implicit functional equivalent if it exists, as in *oświadczenie o rozwiązaniu umowy o pracę przez pracodawcę bez wypowiedzenia*, which may be translated literally as *statement of termination of contract of employment by employer without notice* (as it was in the English translation of the Polish Civil Code) or with a more implicit (but still accurate) equivalent — *notice of summary dismissal*⁹. Without large-scale research it is impossible to determine whether this phenomenon is statistically significant. And last but not least, systematic parallel corpus analysis may help answer a more fundamental question: “Is legal translation more prone to explicitation than other domains?”

4.2. Simplification, disambiguation and avoidance of repetitions

Research on the next universal, simplification, has yielded conflicting results (Mauranen 2006: 95). Simplification is the translator’s inclination “to simplify the language used in translation”, which may be manifested in division of long sentences into shorter ones; avoidance of repetitions; disambiguation; clarifying use of punctuation; lower lexical density, i.e. the ratio of lexical to grammatical words due to a higher frequency of grammatical words; and a lower type-token ratio due to less varied vocabulary compared to non-translated texts (Baker 1996: 181-183). From the perspective of legal translation, two manifestations of simplification may be of interest: avoidance of repetitions and disambiguation. It may be hypothesised that simplification will be found in translation from common law languages (e.g. UK and US English), which favour **repetitions**, such as doublets and synonym strings. Repetitions have served rhetorical purposes due to “the tradition of verbal magic” and are connected with the historic development of legal English under the influence of Latin and French (Mattila 2006: 233): *I hereby give, devise and bequeath to my beloved wife Mary all my estate; This agreement is made, executed and entered into this 17th day of March, 2009 by and between; the decree is final and binding*. The continental system avoids synonymy in legal language; hence, good translation should eliminate any unnecessary repetitions (Gizbert-Studnicki 2001: 45), otherwise a redundant translation would be confusing to a continental lawyer (Mattila 2006: 234). These synonym strings would be translated into Polish as a single word. What is also interesting is to what extent such repetitions (re) appear in the other direction, in translation from a civil-law to common-law system. Would the Polish *testament* be translated as *will* or rather as *last will and testament* or perhaps calqued under the undue SL influence as *testament*? All the three equivalents may be found in translation practice. Another universal, normalisation, hypothesises that translators tend to “exaggerate di-

8 See House for findings on lack of explicitation in German to English translations in contrast to English to German translations (2008: 12).

9 Empirical studies show a failure to lexicalise, i.e. to use a functional equivalent, and a tendency to use more explicit equivalents both in novice and professional translators (see Tirkkonen-Condit 2004: 182).

stinctive features of the target language and to conform to its typical patterns” (Baker 1996: 183) to ensure that it is natural and idiomatic. This may lead to reappearance of repetitions in the other direction, which would not support the hypothesis. For these reasons the avoidance of repetitions merits more attention.

As pointed out by Baker, simplification is connected with explicitation: “simplification involves making things easier for the reader (but not necessarily more explicit), but it does tend to involve also selecting an interpretation and blocking other interpretations, and in this sense it raises the level of explicitness by resolving ambiguity” (1996: 182). Does disambiguation appear in legal translation, and if so, is its extent markedly smaller? While legal language is a compromise between precision and flexibility, ambiguity and vagueness may be strategic and deliberate. For example, the flexible language of legislation allows it “to adapt to differing circumstances and communities (...), to deal with novel situations that are certain to arise in the future as well as changing norms and standards” (Tiersma 1999: 80). A case in point concerns modifiers, such as *reasonable, proper, due*; they become part of legal terms to make their boundaries fuzzy. In the multilingual EU legislation vague language may reflect a compromise or political disagreement between the Member States. If a provision is unclear, it is referred by a national court to the ECJ for interpretation under the preliminary ruling procedure. Disambiguation is perceived as overstepping one’s authority as a translator since translators are expected to retain the same degree of ambiguity and leave disambiguation to courts: “it is generally agreed that the translator has no authority to resolve an ambiguity in the source text as this would be an act of interpretation” (Šarčević 1997: 92). As noted by Northcott/Brown, translators may resolve ambiguities in the ST inadvertently if they lack legal knowledge:

One problem for the translator can be in misunderstanding the deliberate intention to retain ambiguity which can lead to an attempt to make the term more precise and limit possible interpretations by the court. Translators have no authority to resolve ambiguities in source texts. However, this can be brought about inadvertently if translators do not have sufficient legal and linguistic expertise. (2006: 362)

Yet it is not only ignorance but also polysemy which disambiguates. This has been demonstrated by a study of *adverse* on a parallel legal corpus (Goeffroy-Skuce 1997). Polysemy is a natural phenomenon in language: words form networks of interrelated senses (Langacker 1988: 51). Since such networks differ across languages, a translator may have no choice but to pick a single sense and disambiguate in the TL. Again, the scale of disambiguation, whether unavoidable or avoidable, is unknown.

4.3. Textual fit: over-representation, under-representation and untypical collocations

This group of universals is connected with a hitherto neglected aspect of legal translation — the **textual fit**, that is how the translated language (translationese) differs from the non-translated language. Research on comparable corpora suggests that translated texts show the ‘levelling out tendency’ in that they “gravitate towards the centre of a continuum. [...] It involves steering a middle course between any two extremes, converging towards the centre” (Baker 1996: 184). For example, lexical density, type-token ratio and mean sentence length are similar in translated texts, while in non-translated ones they have higher variance (Baker 1996: 184). Furthermore, the distribution of linguistic features may be marked in translations compared to non-translated language: high-frequency lexical items which are shared by the SL and the TL tend to be over-represented, while lexical items which do not have straightforward counterparts in the SL, the so-called unique items, tend to be under-represented (Tirkkonen-Condit 2004: 177). Unique items may be lexical, phraseological, syntactic or textual, and “they do not readily suggest themselves as translation equivalents, as there is no obvious linguistic stimulus for them in the source text” (Tirkkonen-Condit 2004: 177-178). This is closely related to Mauranen’s untypical collocations hypothesis: her research shows that translated language is marked by collocations and colligations which are

possible but rare in the TL and has few combinations which are frequent in the spontaneous TL (2006: 97).

The textual fit hypotheses have not yet been tested extensively, in particular on legal translation. In fact, little is known to what extent translated legal texts differ from spontaneous texts. Nevertheless, these hypotheses raise some challenging questions concerning reception of legal translation by readers. How do such differences affect the readability of translation and its comprehension? To what extent is the marked language which draws attention to itself distracting to readers? Can ‘unnatural’ language of translation adversely affect readers’ attitude to translation? This information may be of relevance to international political constructs founded on legal translations, such as the EU, where translations have met with a cold reception in some Member States, partly due to low quality and partly due to low tolerance of the foreign. The latter applies in particular to ‘minority’ cultures with low exposure to non-native varieties of their local languages, as was the case with the Polish version of the *Acquis Communautaire*, described by the press as “an avalanche of gobbledygook translations”. The more fundamental question, however, is what translators can do to increase the fit and reduce the impression of translationese. The usefulness of such knowledge in training is stressed by Chesterman: “the more we know about T-universals, for instance, the more scholars or trainers will see them as undesirable features that should be avoided – at least in translations whose skopos includes optimum naturalness” (2004: 11).

Given the intensive streams of translation in the globalised world, if translation has an untypical distribution of TL features, it may induce changes in vernacular legal languages. It is acknowledged that “translation is an important way of renewing the target language” (Johansson 2007: 62), and knowledge of the textual fit may help to project future changes. In particular, it seems important for EU official languages which are not working languages of the Union and which are currently experiencing an unprecedented inflow of legal translations into their legislation. Initial research has already been carried out on how the language of Finnish legislation is changing under translated EU legislation (see Piehl 2006); however, more corpus-based studies are needed that will involve other languages.

Research on translation universals and patterns in translated language is still at an early stage and it remains to be seen where it will take us. While S-universals focus on the relationship between the ST and the TT, which is of priority for legal translation, T-universals draw attention to the neglected area of textual fit, which is crucial for translation reception and acceptability. Yet translation universals are not the only object of research on corpora. Equally important is systematic description of the actual translation practice in various genres and language pairs. There are many areas of legal translation that require research on parallel corpora, to name a few: strategies and techniques of dealing with incongruous terms, established equivalents, expert-to-expert versus expert-to-lay communication, and, last but not least, inverse legal translation, which tends to be excluded from corpus-based studies. According to Baker, the most significant contribution of parallel corpora is “a shift of emphasis from prescription to description”, thanks to the possibility of objective observation of solutions used by translators (1995: 231).

5. Practical potential: applicability to legal translator training and bilingual lexicography

In addition to the methodological potential of corpus-based translation studies, corpora have practical applications in translator training, translation practice, lexicography/terminography, and development of information extraction software.

Similarly to the way in which monolingual corpora have revolutionised lexicography, parallel corpora offer new vistas to bilingual lexicography and terminography. Since parallel corpora reflect translation practice and contain “many relevant translation units and their equivalents that tend to be overlooked by lexicographers not working with a parallel corpus (and that is the majority)”, they are used to complement and validate existing dictionaries and reduce arbitrary and

idiosyncratic decisions concerning entry selection (Teubert 2002: 2004), with special extraction software accelerating terminologists' work. Another consequence of corpora is the growing interest in phraseology and impact on equivalence: "(c)orpora have perhaps strengthened the trend away from word-equivalence to phrasal equivalence" (Krishnamurthy 2006: 253). This revolution has not yet fully spread into legal dictionaries, at least not English-Polish ones, which need to be reviewed against parallel corpora and incorporate phraseological elements to be more functional to legal translators. With time electronic bilingual databases should replace traditional printed dictionaries as they will provide contextualised equivalents at a faster speed and will offer new functionalities, such as hyperlinks. Parallel corpora are commonly used in translation practice in the form of CAT tools, such as SDL Trados, Wordfast, DéjàVu, StarTransit, and OmegaT, which store translation memories with aligned SL and TL segments and may be integrated with terminology management software. However, legal translation memories have a limited potential to 'travel' due to confidentiality restrictions. In the future CAT tools may be expected to be integrated with new-generation electronic dictionaries based on parallel corpora.

Apart from professional applications, corpora have pedagogical potential in the translation classroom. As already demonstrated, knowledge of the translation process, translation universals and translationese has straightforward applications in training. In general, corpora are associated with an inductive teaching methodology, referred to as data-driven learning (see Granger 2003: 24) or data-driven discovery learning (see Stubbs 2004: 107). While monolingual and comparable corpora may raise awareness of TL conventions and help students to learn how lawyers write, parallel corpora allow trainees to study translation techniques, established equivalents and recurrent solutions proposed by various professional translators rather than to rely solely on the instructor's experience and intuition. Examples of how legal parallel corpora may be used in translator training have been discussed by Monzó, who reports trainees' increased involvement, productivity and confidence (2008). However, parallel corpora should be applied with care at the beginner level to ensure that trainees do not receive too many ready-made solutions at the expense of reasoned decisions.

All in all, our knowledge of legal translation is still limited. Of increasing importance, legal translation should be researched from as many angles as possible, since each methodology has a unique focus and capacity to look at a different facet of language. The corpus-based approach is one such promising methodology, with a plethora of theoretical and practical applications. It opens up new perspectives on legal language and legal translation by prompting us to ask new research questions. In particular, these questions concern the traditional relation of translations to their STs, as well as the under-researched relation of translations to non-translated language, which may help to assess their potential impact on vernacular legal languages and improve the naturalness of translation by minimising the effects of translation universals and SL interference. If representative and balanced parallel corpora are built in the future, the combination of quantitative and qualitative analysis may lead to new data-driven generalisations on legal translation and contribute to the ongoing discussion within translation studies about the nature of the translation process itself.

References

- Alcaraz Varó, Enrique/Hughes, Brian 2002: *Legal Translation Explained*. Manchester: St Jerome.
- Baker, Mona 1996: Corpus-based translation studies: the challenges that lie ahead. In Somers, Harold L. (ed.), *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamins, 175-186.
- Baker, Mona 1995: Corpora in translation studies: an overview and some suggestions for future research. In *Target* 7(2), 223-243.
- Baker, Mona 1993: Corpus Linguistics and Translation Studies. Implications and Applications. In Baker, Mona/Francis, Gill/Tognini-Bonelli, Elena (eds.), *Text and Technology: In honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 233-250.

- Bhatia, Vijay K./Langton, Nicola M./Lung, Jane 2004: Legal discourse: Opportunities and threats for corpus linguistics. In Connor, Ulla/Upton, Thomas A. (eds.), *Discourse in the Professions. Perspectives from Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins, 203-231.
- Biel, Lucja 2009: Organization of background knowledge structures in legal language and related translation problems. In *Comparative Legilinguistics. International Journal for Legal Communication* 1, 176-189.
- Biel, Lucja 2008: Legal terminology in translation practice: dictionaries, googling or discussion forums? In *SKASE Journal of Translation and Interpretation* [online] 3(1). www.skase.sk/Volumes/JTI03/pdf_doc/BielLucja.pdf (accessed on 28 May 2009).
- Blackwell, Sue 2009: Why forensic linguistics needs corpus linguistics. In *Comparative Legilinguistics. International Journal for Legal Communication* 1, 5-19.
- Blum-Kulka, Shoshana 1986: Shifts of cohesion and coherence in translation. In House, Juliane/Blum-Kulka, Shoshana (eds.), *Interlingual and Intercultural Communication*. Tübingen, Günter Narr Verlag, 17-35.
- Cao, Deborah 2007: *Translating Law*. Clevedon: Multilingual Matters Ltd.
- Carvalho, Luciana 2007: Translating contracts and agreements from a corpus linguistics perspective. In Kredens, Krzysztof/Goźdz-Roszkowski, Stanisław (eds.), *Language and the Law: International Outlooks*. Frankfurt am Main: Peter Lang, 109-121.
- Chesterman, Andrew 2004: Hypotheses about translation universals. In Hansen, Gyde/Malmkjær, Kirsten/Gile, Daniel (eds.), *Claims, Changes and Challenges in Translation Studies. Selected contributions from the EST Congress, Copenhagen 2001*. Amsterdam/Philadelphia: John Benjamins, 1-13.
- Coulthard, Malcolm/Johnson, Alison 2007: *An Introduction to Forensic Linguistics. Language in Evidence*. London/New York: Routledge.
- Engberg, Jan 2002: Legal meaning assumptions – what are the consequences for legal interpretation and legal translation? In *International Journal of the Semiotics of Law* 15(4), 375-388.
- Frawley, William 1984: Prolegomenon to a theory of translation. In Frawley, William (ed.), *Translation: Literary, Linguistic and Philosophical Perspectives*. London/Toronto: Associated University Presses, 159-175.
- Gizbert-Studnicki, Tomasz 2001: Problem przekładu tekstów prawnych. In *Dziedzictwo prawne XX wieku. Księga pamiątkowa z okazji 150-lecia TBSP UJ*. Zakamycze: Kraków, 41-55.
- Goeffroy-Skuce, Anne 1997: Polysemous adjectives in legal translation. In Simms, Karl (ed.), *Translating Sensitive Texts: Linguistics Aspects*. Amsterdam: Rodopi, 155-168.
- Gotti, Maurizio 2001: Semantic and pragmatic values of *shall* and *will* in Early Modern English statutes. In: Gotti, Maurizio/Dossena, Marina (eds.), *Modality in Specialized Texts*. Bern: Peter Lang, 89-111.
- Goźdz-Roszkowski, Stanisław 2006: Frequent phraseology in contractual instruments: a corpus-based study. In Gotti, Maurizio/Giannoni, Davide S. (eds.), *New Trends in Specialized Discourse Analysis*. Bern: Peter Lang, 147-162.
- Granger, Sylviane 2003: The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies? In Granger, Sylviane/Lerot, Jacques/Petch-Tyson, Stephanie (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam/New York: Rodopi, 17-29.
- Gries, Stefan Th. 2006: Introduction. *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*, In Gries, Stefan Th./Stefanowitsch, Anatol. Berlin/New York: Mouton de Gruyter, 1-7.
- Harvey, Malcolm 2002: What's so special about legal translation? In *Meta* [online] 47(2). www.erudit.org/revue/meta/2002/v47/n2/008007ar.pdf (accessed 5 May 2009).
- Heffer, Chris 2005: *The Language of Jury Trial. A Corpus-Aided Analysis of Legal-Lay Discourse*. Basingstoke: Palgrave Macmillan.
- House, Juliane 2008: Beyond Intervention: Universals in Translation? In *trans-kom* [online] 1(1), 6-19. www.trans-kom.eu (accessed 5 May 2009).
- Johansson, Stig 2007: Seeing through multilingual corpora. In Facchinetti, Roberta (ed.), *Corpus Linguistics 25 Years on. Language and Computers*. Amsterdam/New York: Rodopi, 51-71.
- Kenny, Dorothy 2001: *Lexis and creativity in translation. A corpus-based study*. Manchester: St Jerome Publishing.
- Kierzkowska, Danuta (ed.) 2005: *Kodeks tłumacza przysięgłego z komentarzem*. Warszawa: Tepis.
- Krishnamurthy, Ramesh 2006: Corpus Lexicography. In Brown, Keith (ed.), *Encyclopedia of Language and Linguistics*. Vol. 3. Oxford: Elsevier, 250-251.
- Langacker, Ronald W. 1988: A View of Linguistic Semantics. In Rudzka-Ostyn, Brygida (ed.), *Topics in Cognitive Linguistics*. Amsterdam/Philadelphia: John Benjamins, 45-90.
- Mattila, Heikki E.S. 2006: *Comparative Legal Linguistics*. Aldershot: Ashgate.

- Mauranen, Anna 2006: Translation universals. In Brown, Keith (ed.), *Encyclopedia of Language and Linguistics*. Vol. 13. Oxford: Elsevier, 93-100.
- McEnergy, Tony/Xiao, Richard/Tono, Yukio 2006: *Corpus-Based Language Studies. An Advanced Resource Book*. London/New York: Routledge.
- Monzó, Esther 2008: Corpus-based activities in legal translator training. In *The Interpreter and Translator Trainer* 2(2), 221-251.
- Northcott, Jill/Brown, Gillian 2006: Legal translator training: Partnership between teachers of English for legal purposes and legal specialists. In *English for Specific Purposes* 25(3), 358–375.
- Olohan, Maeve 2004: *Introducing Corpora in Translation Studies*. London/New York: Routledge.
- Piehl, Aino 2006: The influence of EU legislation on Finnish Legal Discourse. In Gotti, Maurizio/Giannoni, Davide S. (eds.), *New Trends in Specialized Discourse Analysis*. Bern: Peter Lang, 183-194.
- Pym, Anthony 2004: Localization from the Perspective of Translation Studies: Overlaps in the Digital Divide? In *ELDA* [online]. <http://www.elda.org/en/proj/scalla/SCALLA2004/Pymv2.pdf> (accessed 10 May 2009).
- Šarčević, Susan 2000: Legal Translation and Translation Theory: a Receiver-Oriented Approach. In *Genève 2000: Actes* [online]. www.tradulex.org (accessed 5 May 2009).
- Šarčević, Susan 1997: *New Approach to Legal Translation*. The Hague: Kluwer Law International.
- Stubbs, Michael 2004: Language Corpora. In Davies, Alan/Elder, Catherine (eds.), *Handbook of Applied Linguistics*. Oxford: Blackwell, 106-132.
- Svartvik, Jan 2007: Corpus linguistics 25+ years on. In Facchinetti, Roberta (ed.), *Corpus Linguistics 25 Years on. Language and Computers*. Amsterdam: Rodopi, 11-25.
- Taylor, John 2006: Cognitive Semantics. In Brown, Keith (ed.), *Encyclopedia of Language and Linguistics*. Vol. 3. Oxford: Elsevier, 569-582.
- Teubert, Wolfgang 2002: The role of parallel corpora in translation and multilingual lexicography. In Atenberg, Bengt/Granger, Sylviane (eds.), *Lexis in Contrast*. Amsterdam/Philadelphia: John Benjamins, 189-214.
- Tiersma, Peter M. 1999: *Legal Language*. Chicago/London: University of Chicago Press.
- Tirkkonen-Condit, Sonja 2004: Unique items — over- or under-represented in translated language? In Mauranen, Anna/Kujamäki, Pekka (eds.), *Translation Universals. Do They Exist?* Amsterdam/Philadelphia: John Benjamins, 177-184.
- Tognini-Bonelli, Elena 2001: *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.
- Zanettin, Frederico 2000: Parallel Corpora in Translation Studies. Issues in Corpus Design and Analysis. In Olohan, Maeve (ed.), *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester: St Jerome, 105-118.