

Volume 12/2012

Comparative Legilinguistics

International Journal for Legal
Communication

Institute of Linguistics
Faculty of Modern Languages and Literature
Adam Mickiewicz University
Poznań, Poland

**INSTITUTE OF LINGUISTICS
LABORATORY OF LEGILINGUISTICS**

www.lingualegis.amu.edu.pl

lingua.legis@gmail.com

KOMITET REDAKCYJNY/EDITORIAL BOARD

Editor-in-chief: Aleksandra Matulewska

Co-editor: Joanna Nowak-Michalska

Sekretarze/Assistants: Swietlana Gaś, Karolina Gortych-Michalak, Joanna Grzybek

External Members of the Editorial Board

Lucja Biel, PhD, University of Gdańsk, Poland

Christos Bintoudis, PhD, Greece/Poland

Susan Blackwell, PhD, University of Birmingham, UK

Sheng-Jie Chen, PhD, National Taiwan University of Science and Technology, Taiwan

Ewa Kościółkowska-Okońska, PhD, Nicolaus Copernicus University, Toruń, Poland

Professor Larisa Krjukova, Tomsk State University, Russia

Artur Kubacki, PhD, University of Silesia, Katowice, Poland

Professor Lavinia Nădrag, Ovidius University, Constanța, Romania

Professor Natalja Netšunajeva, Public Service Academy, Estonia

Kiriakos Papoulidis, PhD, Greece/Poland

Professor Fernando Prieto Ramos, ETI, Université de Genève, Switzerland

Professor Judith Rosenhouse, Swantech - Sound Waves Analysis and Technologies, Ltd. (before retirement: Technion I.I.T., Haifa, Israel)

Section editor: Karolina Kaczmarek

Linguistic editors: Jamila Oueslati for Arabic and French, Ya Nan Zhang for Chinese, Colin Phillips for English, Karsten Dahlmans for German, Agnieszka Domagała and Piotr Wierzechon for Polish, Swietlana Gaś for Russian, Eva María Rufo Sánchez-Román for Spanish

Technical editor: Daria Zozula

Adres Redakcji/Editorial Office

Instytut Językoznawstwa

Pracownia Legilingwistyki

Al. Niepodległości 4, pok. 218B

61-874 Poznań, Poland

lingua.legis@gmail.com

Wydanie publikacji dofinansował Instytut Językoznawstwa

Czasopismo znajduje się na liście Ministerialnej czasopism punktowanych z 2012 roku z liczbą 8 punktów.

The issue has been published with financial grant from the Institute of Linguistics, Poland.

Copyright by Institute of Linguistics

Printed in Poland

ISSN 2080-5926

Nakład 100 Egz.

Redakcja i skład: Pracownia Legilingwistyki

Druk: Zakład Graficzny Uniwersytetu im. A. Mickiewicza

Table of Contents

| | |
|--|------------|
| Preface | 5 |
| ARTICLES | |
| 1. Legal Language and Terminology | |
| Michal KUBÁNEK, Ondřej KLABAL (Czech Republic) Plain or Archaic: The New Czech Civil Code Going against the Flow | 9 |
| Mami Hiraike OKAWARA (Japan) Lay Understanding of Legal Terminology in the Era of the Japanese Lay Judge System | 19 |
| 2. Legal Translation | |
| Hanna JUSZKIEWICZ (Poland) Dissolution of Marriage: Functional Approach to Polish-English Translation of Selected Court Documents | 49 |
| Łukasz ZYGMUNT (Poland) Lexical Pitfalls in Polish-English Legal Translation: a Case Study Involving Students of English Philology in Poland | 67 |
| 3. Corpora | |
| Juliette SCOTT (UK) Can Genre-Specific DIY Corpora, Compiled by Legal Translators Themselves, Assist Them in 'Learning the Lingo' of Legal Subgenres? | 87 |
| 4. Reviews | |
| Marcus GALDIA <i>An Appraisal of Language and Law in Times of Expanding Legal Linguistics</i> . A review of <i>The Oxford Handbook of Language and Law</i> edited by Peter M. TIERSMA, Lawrence M. SOLAN, 2012. Oxford University Press. | 101 |
| Karolina KACZMAREK <i>Precision and Vagueness in Legal Linguistics</i> . A review of <i>Vagueness in Normative Texts. Linguistic Insights 23</i> . Studies in Language and Communication. edited by Vijay K. BHATIA, Jan ENGBERG, Maurizio GOTTI, Dorothee HELLER, 2005. Peter Lang. | 113 |
| Aleksandra MATULEWSKA. <i>New Developments in Legal Linguistics</i> . A review of <i>Legal Discourse Across Languages and Cultures. Linguistic Insights 117</i> . Studies in Language and Communication. edited by Maurizio GOTTI, Christopher WILLIAMS, 2010. Peter Lang. | 115 |
| Permanent Advisory Board and Reviewers 2012 | 117 |

Spis treści

| | |
|--|------------|
| Wprowadzenie | 5 |
| ARTYKULY | |
| 1. Język prawa i terminologia prawnicza | 9 |
| Michał KUBÁNEK, Ondřej KLABAL (Czechy) Prosty czy archaiczny: nowy czeski kodeks cywilny idzie pod prąd | |
| Mami Hiraike OKAWARA (Japonia) Rozumienie terminów prawnych przez niespecjalistów w dobie systemu opartego na sędziach niezawodowych w Japonii | 19 |
| 2. Tłumaczenie prawnicze | 49 |
| Hanna JUSZKIEWICZ (Polska) Funkcjonalne podejście do polsko-angielskiego przekładu wybranych pism procesowych i orzeczeń sądowych w sprawach rozwodowych oraz w sprawach o orzeczenie separacji | |
| Łukasz ZYGMUNT (Polska) Leksykalne pułapki w polsko-angielskim przekładzie prawniczym: studium przypadku obejmujące studentów filologii angielskiej | 67 |
| 3. Korpusy | 87 |
| Juliette SCOTT (Wielka Brytania) Czy korpusy poszczególnych gatunków tekstów zebrane samodzielnie przez tłumaczy prawniczych pomagają im „uczyć się języka” podgatunków tekstów prawnych? | |
| 4. Recenzje | 101 |
| Marcus GALDIA <i>Pochwała języka i rawa w czasach rozwoju legilingwistyki</i> . Recenzja monografii pod redakcją Petera M. TIERSMY, Lawrence’a M. SOLANA, <i>The Oxford Handbook of Language and Law</i> . 2012. Oxford University Press. | |
| Karolina KACZMAREK <i>Precyzja i niedookreśloność w legilingwistyce</i> . Recenzja monografii pod redakcją Vijay K. BHATII, Jana ENGBERGA, Maurizio GOTTIEGO, Dorothee HELLER, <i>Vagueness in Normative Texts. Linguistic Insights</i> . Studies in Language and Communication. 2005. Peter Lang. | 111 |
| Aleksandra MATULEWSKA <i>Nowe trendy w legilingwistyce</i> . Recenzja monografii pod redakcją Maurizio GOTTIEGO, Christophera WILLIAMSA, <i>Legal Discourse Across Languages and Cultures. Linguistic Insights 117</i> . Studies in Language and Communication. 2010. Peter Lang. | 113 |
| Komitet redakcyjny, doradcy i recenzenci 2012 | 117 |

CAN GENRE-SPECIFIC DIY CORPORA, COMPILED BY LEGAL TRANSLATORS THEMSELVES, ASSIST THEM IN ‘LEARNING THE LINGO’ OF LEGAL SUBGENRES?

Juliette SCOTT, MA

University of Portsmouth, School of Language and Area Studies,
Park Building, King Henry I Street, Portsmouth PO1 2DZ, UK
juliette.scott@port.ac.uk

Abstract: This paper presents a study aimed at examining whether DIY corpora compiled by professional legal translators can assist them in their role as learners of legal sublanguages, particularly those translators working into non-native target languages.

A procedural DIY corpus methodology has been developed, involving the framed retrieval of authoritative legal texts from Internet repositories or other sources by legal translators themselves, according to their specific needs or those of particular projects, bearing in mind at all times feasibility in the workplace. Target audience expectations and requirements are also an important consideration in the project.

A pilot study performing some initial testing with professional legal translators in certain legal genres and different languages has been completed and will be reported on. Results so far seem to indicate that compilation of such corpora can be achieved in an average of 30-45 minutes, in line with users' expressed criteria.

It is posited that these highly specialised corpora may provide translators with some additional reference material that they are sorely lacking due to the absence or shortage in many language combinations of legal dictionaries or thesauri, in particular as regards collocations. It is hoped that a contribution may be made to professional practice in the long term.

CZY KORPUSY POSZCZEGÓLNYCH GATUNKÓW TEKSTÓW ZEBRANE SAMODZIELNIE PRZEZ TŁUMACZY PRAWNICZYCH POMAGAJĄ IM „UCZYĆ SIĘ JĘZYKA” PODGATUNKÓW TEKSTÓW PRAWNYCH?

Artykuł przedstawia wyniki badań przeprowadzonych w celu określenia czy korpusy tekstów tworzone przez zawodowych tłumaczy testów prawa, zwłaszcza tych tłumaczących na języki obce, mogą pomagać im w uczeniu się odmian języka prawa. Stworzono metodologię dotyczącą samodzielnego tworzenia korpusu wykorzystującego pozyskiwanie wiarygodnych tekstów z internetu lub innych źródeł przez tłumaczy. Badanie pilotażowe dotyczące wybranych gatunków tekstów prawa oraz różnych języków dowodzi, że kompilacja takiego korpusu zgodnego z kryteriami użytkowników, może zająć 30-45 minut. Zakłada się, że te wysoko wyspecjalizowane korpusy stanowią dla tłumaczy dodatkowe źródło informacji, zwłaszcza, że w wielu językach brakuje słowników prawnych, szczególnie jeśli chodzi o kolokacje.

Introduction: Scope and limitations

Legal translation²⁷ is a thorny subject, due in part to the highly conservative nature of legal sublanguages. Whilst lawyers are taught legal language at law school as part of their studies, and paralegal staff undertake training on legal writing, translators are rarely trained in this singular lect.

Despite the fact that professional legal translators have a wide range of *general* tools at their disposal, such as dictionaries, glossaries, termbases, and online fora²⁸, target sublanguage conventions and appropriate collocations may escape them due to the lack of legal thesauri²⁹ and legal collocation dictionaries (Biel 2010). Parallel corpora, generally used by professionals in the form of translation memories, may make a contribution, but availability is limited for the most part to the genre of legislation.

The project to be described in this paper seeks to examine whether and to what extent small, specialised electronic corpora created in legal subgenres by professional translators themselves according to their needs can assist them in producing translations that are closer in line with target audience expectations.

The project also aims, in a distinct strand, not detailed here, to gather information from those commissioning legal translations in order to better understand their expectations, as well as requirements and quality issues encountered, and to find out what information is communicated to translators when commissioning takes place.

The majority of studies on corpora created by translators themselves (hereafter ‘DIY corpora’, standing for Do-It-Yourself) have involved students. This project will involve professional translators only. Freelancers have been selected rather than translators employed in-house, because the latter generally have access to far more terminological resources to support their efforts.

Research methods

This empirical research project is mainly qualitative – participants are volunteers and not selected using statistical methods. However, insofar as certain demographic data such as country of origin, language pair, professional experience, etc. is being collected, the project could also be described as adopting a mixed methods approach. It comprises a pilot study performed between December 2011 and March 2012, and a main study to be launched in October 2012. The author has herself been a practising professional legal translator for over twenty years, and there is thus an element of action research.

Recruitment of participants so far has been carried out using web groups such as Yahoo! Groups and LinkedIn, through professional blogs, and face-to-face networking at events on translation and the law. The main study may also involve workshop settings – this has been suggested by certain interested national associations³⁰.

Feedback is obtained through questionnaires hosted on an online platform – Wufoo.com. This enables the researcher to have real-time access to the data as it is

²⁷ This paper is concerned with translation (the written medium) and not interpreting (spoken).

²⁸ Such as Proz.com, Wordreference.com.

²⁹ Burton’s Legal Thesaurus published by McGraw Hill, in English, would appear to be the only example at this time.

³⁰ E.g. Sweden, Holland.

generated, and to produce statistics on an ad hoc basis. Data from the Wufoo platform can be exported, via Microsoft Excel, to the data analysis software package NVivo, in particular for responses to open questions, where text can be marked up into fields created by the researcher on an on-going basis and subsequently enable a fine-grained scrutiny. It is hoped that these tools will give a more rigorous justification to qualitative findings, and provide deeper insights into the data.

Disciplinary foundations

Genre theory applied to the legal domain

In this paper I shall use the following terms for the legal domain: *supergenre* as comprising all “language of the law” (Bhatia 1987); in which ‘*subgenres*’ are grouped under ‘*genres*’, which act thus as common headings encompassing a number of the lower level categories” (Monzó Nebot 2008, emphasis added).

Bhatia (1987, 227) provided a structure distinguishing the main legal genres according to their “communicative purposes” and the “lexico-grammatical, semantico-pragmatic, and discoursal resources” used in different legal contexts. A systematic organisation of such a complex supergenre is clearly valuable. However, regarding written genres, Bhatia (1987) differentiated between “frozen legal documents like contracts, agreements, insurance policies, etc.” and “formal” documents such as “legislation, rules and regulations, etc.”; classifying them both under the term “legislative”, which seems somewhat perplexing. In his 2006 work, however, Bhatia offered a different view of legal genres, dividing them into a “primary” genre – legislation; “secondary” genres – e.g. judgments and case reports; “enabling academic genres” – such as textbooks, critical essays, etc.; and “target genres” – e.g. contracts, affidavits, insurance documents etc. The latter three categories are collectively termed “derived” genres, as being “interpretations of legislative intentions” (p. 6).

Other classifications of the legal supergenre have been proposed, for example by Trosborg (1997, 20) according to “situation of use”. Kurzon (1997, 120) differentiates between “legal language [...] employed when people talk about the law” and “the language of the law” that he defines as institutional laying down of the law. A further classification has been made by Mattila (2006, 4–5) according to the sub-groups of legal professionals.

I will be offering a complementary structure aimed at addressing use within legal translation practice. In particular, such an organisation could be used by translators to classify their collection of DIY corpora.

Legal translation and functionalism

The issue of a legal translation’s (non-)compliance with target expectations will be explored in this section.

In a work devoted to the interaction between translation theory and practice, Chesterman and Wagner (2002) raise three key issues: differentiation of translators’ service by offering several levels of quality to suit a range of purposes; “maintaining a resemblance to the intended genre” (e.g. p. 96) when translating; and increasing the use of target language resources in the profession.

Nord (1997) is a seminal work on functionalist approaches to translation. It contains a detailed description of *Skopos*theorie, according to which “the prime principle

determining any translation process is the purpose (*Skopos*) of the overall translational action” (p. 27). Nord also emphasises the importance of a translation ‘brief’ in translator training, defining the “conditions under which the target text should carry out its particular function” (pp. 59–60).

In the legal domain, Šarčević (2000) outlined the new focus placed by Reiss and other German scholars from the 1970s onwards on target audiences, to enable translators to select appropriate translation strategies, discussing *Skopostheorie* in particular. She also considers differentiation of purpose, and the corresponding variation in strategies. She cites Kelsen in classifying different groups of addressees or receivers as direct or indirect (p. 4), and further develops approaches and advice for specific receiver groups.

Gémar (2002, 168) distinguished four main categories of reader in the context of legal texts: laymen, those who are ‘lettered’, practising legal professionals, and legal scholars. It ensues that depending on its destination, a translation will be informed by knowledge of its intended readership.

Garzone (2000) demonstrated the suitability of a functional approach to legal translation and concluded that:

“the degree of equivalence to be achieved in the translation of a given text is not absolute, but depends first and foremost on the TT [target text] intended function as well as on the nature of the ST [source text]; the whole process is governed by a principle located at a sufficiently high level of generalisation as to be suitable for virtually all types of legal texts”.

Sparer and Covacs stated that “Specialists in legal translation now define fidelity as achieving an equivalent impact on the target reader [...] to respect the stylistic conventions of the target legal culture” (taken from Harvey, 2002, 180).

However, in order for the translation to appropriately fulfil the intended purpose, it should be clear from the above that translators need to be informed of what that purpose actually is. The latter point is beyond the scope of this paper, but as mentioned in the Introduction, forms part of the second strand of the project.

Corpora and translation

Although corpus use in translation has been studied extensively at academic level since first highlighted by Baker (1993, 1995), Kenny (2001) and Olohan (2004), amongst others, professional uptake is considerably more limited, as demonstrated in the EU-funded MeLLANGE project survey (MeLLANGE, 2006) and in my own survey (J. Scott 2011, 7–8) and discussed by Bowker (2004) and Bernardini (2006).

Much research has been done on parallel and comparable³¹ corpora, and a significant amount on small DIY corpora *per se* (e.g. Varantola 2002; Zanettin 2002; Maia 2002).

³¹ The terms ‘parallel’ and ‘comparable’ are somewhat contentious in the literature. They are considered here as defined by Fernandes (2006): ‘parallel’ being texts and their translations, ‘comparable’ consisting of two (or more) sets of L1 texts.

In particular Varantola (2002) described “disposable corpora as intelligent tools” for translators, i.e. those that “adapt to users’ needs and allow user profiling” (p. 171). She also states that “disposable material can be recycled and refined to form part of a more permanent collection” (p. 175).

Other studies have examined DIY corpora used in specific genres such as tourism (Wilkinson, 2007) and timber (Jääskeläinen & Mauranen: 2006), and some cover several genres (Tagnin 2002). However, studies have in the main concerned student rather than professional translators – indeed my searches in this regard have only brought to light Jääskeläinen & Mauranen (2006) and Maher, Waller & Kerans (2008), and there has been little application to the legal field – there is only a passing reference in the latter paper to an association bylaws corpus. There is a study involving a corpus of travel insurance documents and legislation, in the context of the extensive Spanish tourist industry rather than legal translation as such, mainly focusing on translator training and corpus representativeness (Corpas Pastor & Seghiri 2009).

Corpora and Language for Special Purposes

Bowker and Pearson (2002) is generally considered as the reference work on the compilation and use of corpora in the field of Language for Special Purposes (LSP). They state: “In our experience, well-designed corpora that are anywhere from about ten thousand to several hundreds of thousands of words in size have proved to be exceptionally useful in LSP studies”.

The use of self-compiled specialised corpora for LSP learning was studied by Lee and Swales (2006) in the area of English for Academic Purposes (EAP). A group of university students were introduced to the corpora and to the skills needed to investigate the language, *inter alia* using context for disambiguation, using frequency patterns, and examining collocations. They were also taught to compile two corpora of their own. The software used was WordSmith Tools (M. Scott 1996). Upon completion of the research programme, most participants had purchased the software in order to “continue their concordancing activities outside the classroom”.

Hafner and Candlin (2007) examined the use of specialised corpora in understanding professional discourse. Their study, carried out in Hong Kong, looked at how corpus tools could contribute to language prowess as a part of professional training for fledgling lawyers. A genre-based approach was adopted and examples included: letters of advice; statements of claim, defence and counterclaim; affidavits; opinions; and agreements. Student participants were given access to a corpus of 114 legal cases (797,000 words) from three subject areas and divided into seven sub-corpora, that had been compiled for them, and were given the option whether to use the tool or not. Those who chose to adopt the corpus “viewed them as a convenient and helpful toolset to provide language support for their legal writing and drafting assignments” (p. 314).

Bhatia, Langton and Lung (2004), looking at corpus linguistics and language teaching and learning in legal contexts, conclude, citing Tribble, *inter alia*, that “the use of genre-based small corpora will be much more useful than large corpora covering a complete register of law” (2004, 215). They further hold that “legal discourse is so conservative in its construction, interpretation and use that it often does not require a large corpus to determine its linguistic frequencies” (2004, 207).

The NIFTY corpus methodology

I will now describe the corpus methodology that has been developed. It is a process consisting of five steps: defining corpus criteria; collecting corpus material; manual assessment; file conversion³²; and use in translation. The current version of the methodology results from trialling during a pre-study that I carried out from 2009–2011 which is the subject of a Masters' thesis (J. Scott 2011).

I have chosen the acronym 'NIFTY', standing for 'Nimble, Individual, Fast & fruitful, Tailor-made and Yield a great harvest!', as well as the usual dictionary definitions of the word³³. The aim is to make the methodology as user-friendly as possible.

Defining corpus criteria

In particular: target language; geographical perimeter, due to differences between UK/US English or Portuguese in Brazil and Portugal, for example; type of document [the (sub)genre]; file type, since the PDF format may be more freely available than .doc files in the legal domain and may show signatures or official stamps thus attesting to the authoritative nature of the text; and date of publication, e.g. where terminology has changed following a certain event or piece of legislation.

Collecting corpus material

As regards corpus collection, translators may make use of their own archives, and are also given some indications on legal electronic document repositories, both those accessible freely and those on a subscription basis. I have also provided step-by-step instructions for advanced Google file search.

Manual assessment

The importance of a short but essential manual assessment is strongly emphasised, to check whether texts are authoritative, include good quality language (are not obviously translations or badly written), and whether they comply with the criteria defined. This is feasible since the number of texts in a corpus for a highly specific legal subgenre remains manageable³⁴.

Corpus query software and file conversion

Following comparative testing (J. Scott 2011), the freeware AntConc (Anthony 2010) was chosen in order to consult the corpora collected. Preference was given to this software rather than WordSmith Tools (M. Scott, 2010) due to its single user interface and simplicity of use. AntConc is also freeware, whereas WordSmith Tools costs GBP 59.50³⁵. At the present time, corpus query software can only accept .txt files. Some

³² Due to restrictions in corpus query software packages at this time.

³³ Examples: first-rate; great, clever, sharp: a nifty idea.

³⁴ Testing carried out by the author so far of corpora comprising an average of between 10 to 25 texts has yielded good results. These figures vary according to the genre.

³⁵ Price checked 31 August 2012.

examples of batch file conversion software are therefore supplied, where possible freeware. Using such packages, all files collected can be converted in a single operation. The importance of clear labelling of files is stressed in the participants' training pack, as highlighted by Maher, Waller and Kerans (2008), to ensure that source files can be identified at the corpus consultation stage. A suggested file name might include a language code, type of document (subgenre) and name(s) of party(ies).

Corpus consultation and use during translation

In terms of corpus consultation and use while performing a translation, as regards the corpus query software, the attention of study participants is drawn to the Key Word in Context (KWIC) concordance function, keyword lists, and cluster/collocate functions.

Once familiar with the methodology, participants may also choose to load more than one set of files (NIFTY corpus) into the software, and thus consult several corpora at the same time, where relevant for a translation. For example, if they have collected corpora for power purchase agreements, joint venture agreements and sales agreements, they may choose to consult these three corpora if performing a translation of another type of agreement. In this case, they could avail themselves of the proposed supergenre structure referred to briefly in the section on genre theory.

Pilot study

Participant profiles

Current practice in the translation profession is heavily dependent on the Internet, both for job offers – for example through Proz.com and similar, and email exchanges with regular clients – and to consult resources such as termbases. As described earlier, online groups have therefore been leveraged in order to recruit participants.

This may lead to certain type of 'internet-savvy' participant being foregrounded in the study, but I have weighed this up with the advantages of obtaining global reach and thus a wider range of languages. By the very nature of their working methods as described above, translators should tend towards this profile in any case. Wright (2006) carried out a study of the strengths and weaknesses of researching Internet-based populations and online survey research, in which he concludes that researchers may save considerable time using online survey tools, but must be aware of "issues related to sampling frames, response rates, participant deception, and access to populations" in respect of their research aims.

Data collection

Three questionnaires are being used. A registration form, setting out the appropriate ethical information, is used to collect data relating to the translator's profile. Upon receipt of this form, a participant code is issued, enabling the remaining data to be anonymised. Participants were asked to provide feedback using two separate forms – one relating to corpus compilation, and another on corpus use in translation.

Following registration, professional translators participating in the research are given a training pack consisting of an illustrated handbook as well as a video guide in the form of a commented screencast slide show giving details on how to compile their

NIFTY corpora. These multimedia tools can be downloaded from a simple project website that I created to streamline the administration of participants. The website is password-protected in order to control participation in the study and avoid undue dissemination.

In order to recruit professional translators, to date I have used the following sources:

- i) a pool of 81 translators that had expressed interest in further research during a pre-study (J. Scott, 2011);
- ii) 106 translators that were present at a legal translation conference where I ran an introductory workshop on the methodology;
- iii) a notice posted on a private legal translators' group on Yahoo;
- iv) a notice posted on an open legal translators' group on LinkedIn.

This recruitment drive for the pilot study took place in early January 2012, resulting in 43 translators registering to participate in trialling the methodology.

Results

The following section summarises some key aspects of the data collected so far.

Table 1 shows an attrition rate of 79% between registration and provision of feedback. Many participants cited workload or personal circumstances when contacted on this subject. Some have requested to defer their participation until the main study.

Table 1: Overview of pilot study participants

| | |
|---|-----------------|
| N° of translators registered | 43 |
| N° of target languages registered | 12 |
| N° of countries registered | 16 |
| N° of translators having provided feedback on corpus compilation and/or use | 9 ³⁶ |

Participants were free to choose the content of their NIFTY corpora. It can be seen from Table 2 below that the subgenres chosen range from court-related documents (e.g. pleadings and judgments) to corporate texts (e.g. agreements, Articles of Association), notarial texts (e.g. deeds) and regulatory instruments (e.g. rules, procedures, codes).

³⁶ One of these participants is an academic also working as a translator in an institution and thus does not fit into the intended target population frame.

Table 2: Data for NIFTY corpora compiled by pilot study participants

| (Sub)genre | Total words | Language | Source of corpus material | Time required for compilation (in minutes) | Participant code |
|---|--------------|----------|-----------------------------------|--|------------------|
| Divorce judgments | 5 527 | Spanish | Own archives | 15 | 018 |
| Divorce decrees | Not supplied | Polish | Google file search | Not supplied | 035 |
| Acceptance and Vesting of Estate | 12 500 | Spanish | Own archives | 40 | 018 |
| Lease agreements | 1 643 | English | Web repository | 40 | 036 |
| Loan agreements | 120 000 | English | Online search | 30 | 050 |
| Sales agreements | 45 000 | Spanish | Own archives & Google file search | 30 | 043 |
| Memoranda & Articles of Association | 437 971 | English | Own archives & Google file search | 20 | 029 |
| Articles of Association | 221 283 | English | Google file search | 45 | 031 |
| Deed of acceptance and award of inheritance | 2 000 | Catalan | Own archives & Google file search | 40 | 043 |
| International Arbitration Rules | 115 847 | Spanish | Official websites | 75 | 018 |
| Patent examining procedures | 67 578 | English | Google file search | 60 | 017 |
| Penal codes | 206 089 | Spanish | Online search | 60 | 014 |
| Study and examination codes | 70 000 | English | Online search | 30 | 050 |
| <i>Arrhythmias</i> ³⁷ | 15 734 | English | Not supplied | 60 | 017 |

Feedback received from pilot study participants' structured questionnaires can be summarized in three main points. Firstly, it was generally seen as relatively easy to compile the corpora. Secondly, all participants providing feedback felt that the tool was useful. Several also felt more confident in their completed translation than without the use of the tool. Third, they used the corpora more than they had expected, and used other tools and methods less than usual.

Table 3 below includes comments sent in by participants either by email or using the free comments box in the questionnaire. As can be seen, the comments are favourable to the use of the methodology. The position of those who did not provide feedback is not known. The author is currently considering ways in which provision of feedback could be incentivised in the main study.

³⁷ Whilst this is a specialised corpus, it does not fall within the frame of legal (sub)genres.

Table 3: Pilot study participants' free comments – raw data

| Participant code & country | Comment |
|----------------------------|---|
| 018, Spain | <i>"I really find your proposal very useful. I guess it depends on each one own translating method. In my case, I really used that function, I need to check on original texts (into the documents I really trust) over and over. I used to did it searching manually or with google desktop help. NIFTY corpus... it's a great tool, I am glad that I went to Lisbon and hear you. I am going to save time and, specially, increase the confidence on my own job. ¡Muchas gracias!"</i> |
| 018, Spain | <i>"Just some comments about the collecting process. It's my first corpus, I did it in a field that I usually work with and feel comfortable. To my surprise, I have used the corpus more than I expected. I have used original court texts from my own archives that were in DOC format. I have a lot of original useful PDFs in my own archives, but when I tried to use them, even if I use Abby finereader 9.0, I realised that I had to work for a while on them and I don't have time now. After using the corpus I have realised that it is a great tool and will work on my PDF files to have clean TXT files to work with."</i> |
| 022, Germany | <i>"I've started working through Juliette's material. The introduction video she did is a beautifully simple, clear explanation of basic principles and terms, and after watching it a few things that I thought I knew were understood much better. It also dawned on me after a bit that, while her approach is similar in many ways to what I have done for some time in source language research, she focuses on the target language. This has pretty powerful implications for someone working to master a new specialty. I had never really thought about this much before, because I mostly translate in domains I know very well because I have worked in them at some point as a researcher, etc. and I'm not much bothered about searching for collocations and the like. Her research project involves legal translators. But her methodology applies very well to any specialist domain. And I do see value in it for my usual specialties (including two legal areas), because it is a more efficient way of performing certain kinds of language checks. I've just been so focused on source language that this took a while to sink in."</i> |
| 014, USA | <i>"I just - finally - dared to dive in and create my first corpus... and I feel as if I've just learned the ABC! This may sound too gushy, but I'm truly happy about this. I had read Lynne Bowker's book, but I was still intimidated by the prospect of building a corpus. Now it all worked beautifully, thanks to the excellent instructions on your guide..."</i> |
| 050, Czech Republic | <i>"The bulk of my translation work are translations into English, and I think that as a non-native translator I need much more robust language data to consult in order to achieve natural collocations and expressions typical of the given text type. Therefore, I often work with language corpora (mostly Mark Davies). However, the problem is that these corpora are very often too large on the one hand, and little specialized on the other. That's why I was happy to have been introduced to NIFTY corpora, which I started using almost immediately. I usually download about ten documents to compile the corpus. Corpora I have made include a corpus of study and examination codes, a corpus of loan agreements etc. I have always found them to be very useful, and they have helped me to use idiomatic language structures and collocations. I would especially highlight their benefit for translations into L2."</i> |

These preliminary results, however, need to be analysed in more detail, and further investigated, in all probability by means of interviews.

Data collection challenges

It has proved very difficult to obtain feedback from participants, despite their expressed high levels of interest in the tool. A number of reasons have been cited, the majority relating to heavy workloads. This difficulty has already been encountered in the small number of previous studies involving professional translators (Jääskeläinen and Mauranen, 2006).

At this stage it is unknown whether they are using the tool and not providing feedback, encountering difficulties, or simply not using the tool at all. During the main study attempts will be made to drill down into this issue³⁸.

Forthcoming main study

The main study, to be launched in October 2012, will aim to include a larger number of professional legal translators, and to encompass other languages.

Recruitment will once again use online channels such as LinkedIn groups and professional blogs in order to achieve a wide geographical spread, as well as translator networks. As mentioned in the section on research methods, workshops may also be run in collaboration with national translators' associations.

In order to address the problems of data collection, supplementary contact methods will be made available in addition to the online feedback forms.

Conclusion

This paper has presented initial findings concerning a methodology for monolingual target-language corpora as a tool to counterbalance the lack of familiarity by legal translators with the 'insider' sublanguages of the law, in particular with a view to their usefulness in retrieving legal collocations.

The pilot study has shown some encouraging avenues to be explored, but has underscored the constraints of carrying out research with professional translator participants in terms of their availability.

The next stages of the research will involve a deeper analysis of the data already collected, and in the light of preliminary results, slight changes to the collection instruments and methods. It is hoped that the main study will include a larger participant population, as well as a number of in-depth interviews to further probe how translators use the tool, and to assess its value and practicability in the workplace.

³⁸ In addition, use in other fields than law has been alluded to twice: one corpus was compiled for arrhythmias, and one participant mentioned use "in other areas" (see Table 3).

Bibliography

- Baker, Mona. 1993. Corpus linguistics and translation studies. Implications and applications. In M. Baker, G. Francis and E. Tognini-Bonelli, eds., *Text and Technology: In Honour of John Sinclair*, pp. 233–50. Amsterdam: John Benjamins.
- Baker, Mona. 1995. Corpora in translation studies: An overview and some suggestions for future research, *Target* 7(2), 223–43.
- Bernardini, Silvia. 2006. Corpora for translator education and translation practice Achievements and challenges. *Third International Workshop on Language Resources for Translation Work, Research & Training*, 17–22. Retrieved August 13, 2011 from <http://hmk.ffzg.hr/bibl/lrec2006/workshops/W17/proceedingsLR4TransIIIey.pdf#page=23>
- Bhatia, Vijay, K. 1987. Language of the law. *Language Teaching* 20, 227–234.
- Bhatia, Vijay, K., Langton, Nicola M., & Lung, Jane. 2004. Legal discourse: Opportunities and threats for corpus linguistics. In U. Connor, T. A. Upton, eds, *Discourse in the professions. Perspectives from corpus linguistics* (pp. 203–231). Amsterdam: John Benjamins.
- Biel, Lucja. 2010. The textual fit of legal translations: focus on collocations in translator training. In: Ł. Bogucki, ed., *Teaching Translation and Interpreting: Challenges and Practices*, pp. 25–39. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Bowker, Lynne. 2004. Corpus resources for translators: academic luxury or professional necessity? *TradTerm*, 10, 213–247.
- Bowker, Lynne & Pearson, Jennifer. 2002. *Working with specialized language: a practical guide to using corpora*. London: Routledge.
- Chesterman, Andrew, & Wagner, Emma. 2002. *Can theory help translators?: a dialogue between the ivory tower and the wordface*. Manchester: St Jerome.
- Corpas Pastor, Gloria & Seghiri, Miriam. 2009. Virtual corpora as documentation resources: Translating travel insurance documents (English-Spanish)*. In A. Beeby, P. Rodríguez Inés and P. Sánchez-Gijón, eds., *Corpus use and translating: corpus use for learning to translate and learning corpus use to translate*, pp. 75–107. Amsterdam: John Benjamins.
- Fernandes, Lincoln. 2006. Corpora in translation studies: Revisiting Baker’s typology. *Fragmentos*, 30, 87–95.
- Garzone, Giuliana. 2000. *Legal and functionalist approaches: A contradiction in terms?* Paper presented at Legal translation, history, theory/ies, and practice. Retrieved October 17, 2011 from <http://www.tradulex.org/Actes2000/Garzone.pdf>
- Gémar, Jean-Claude. 2002. Le plus et le moins-disant culturel du texte juridique. Langue, culture et équivalence. *Meta*, 47(2), 163–176.
- Hafner, Christoph, & Candlin, Christopher. 2007. Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes*, 6 (4), 303–318. doi: 10.1016/j.jeap.2007.09.005
- Harvey, Malcolm. 2002. What’s so special about legal translation? *Meta*, 47(2), 177–185.

- Jääskeläinen, Riitta & Mauranen, Anna. 2006. Translators at work: a case study of electronic tools used by translators in industry. In G. Barnbrook, P. Danielsson, and M. Mahlberg, eds, *Meaningful texts: the extraction of semantic information from monolingual and multilingual corpora*, pp. 48–53. London: Continuum International.
- Kenny, Dorothy. 2001. *Lexis and creativity in translation: a corpus-based study*. Manchester: St. Jerome.
- Kurzon, Dennis. 1997. ‘Legal language’: varieties, genres, registers, discourses. *International Journal of Applied Linguistics*, 7(2), 119–139.
- Lee, David & Swales, John. 2006. A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25, 56–75.
- Maher, Ailish, Waller, Stephen & Kerans, Mary E. 2008, July. Acquiring or enhancing a translation specialism: The monolingual corpus-guided approach. *The Journal of Specialised Translation*, 10. Retrieved August 13, 2011 from http://www.jostrans.org/issue10/art_maher.php.
- Maia, Belinda. 2002. Do-it-yourself, disposable, specialised mini corpora – where next? Reflections on teaching translation and terminology through corpora. *Cadernos de Tradução IX - Tradução e Corpora*, 1(9), 221–236. Retrieved August 13, 2011 from <http://www.periodicos.ufsc.br/index.php/traducao/article/view/5987/5691>
- Mattila, Heikki, E.S. 2006. *Comparative legal linguistics*. Aldershot: Ashgate.
- MeLLANGE (Multilingual eLearning in LANGuage Engineering) 2006, April 20. *Corpora & e-learning questionnaire results summary*. Retrieved January 13, 2011 from <http://mellange.eila.univ-paris-diderot.fr/>
- Monzó Nebot, Esther. 2008. Corpus-based activities in legal translator training. *The Interpreter and Translator Trainer*, 2(2), 221–252.
- Nord, Christiane. 1997. *Translating as a purposeful activity*. Manchester: St Jerome.
- Olohan, Maeve. 2004. *Introducing corpora in translation studies*. London: Routledge.
- Šarcevic [sic], Susan. 2000, February 17–19. *Legal translation and translation theory: a receiver-oriented approach*. Paper presented at Legal translation, history, theory/ies, and practice. Retrieved August 29, 2011 from <http://tradulex.org/Actes2000/sarcevic.pdf>
- Scott, Juliette, R. 2011. *DIY corpora: a pearl in the legal translator’s sea of tools*. Unpublished masters dissertation, University of Portsmouth, Portsmouth.
- Tagnin, Stella, E. O. 2002. Corpora and the innocent translator. *inTRAlinea*, 5 (Special issue). Retrieved August 13, 2011 from: http://www.intralinea.it/specials/cult2k/eng_more.php?id=128_0_42_0_M%25
- Trosborg, Anna. 1997. *Rhetorical strategies in legal language: discourse analysis of statutes and contracts*. Tübingen: Gunter Narr Verlag Tübingen.
- Varantola, Krista. 2002. Disposable corpora as intelligent tools in translation. *Cadernos de Tradução IX – Tradução e Corpora*, 1(9), 171–189.
- Wilkinson, Michael. 2007, January. Corpora, serendipity & advanced search techniques. *The Journal of Specialised Translation*, 7. Retrieved August 28, 2011 from http://www.jostrans.org/issue07/art_wilkinson.php

- Wright, Kevin, B. 2006. Researching Internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication* 10(3). doi: 10.1111/j.1083-6101.2005.tb00259.x
- Zanettin, Federico. 2002. Corpora in translation practice. In E. Yuste-Rodrigo, ed., *Language resources for translation work and research, LREC 2002 Workshop Proceedings, Las Palmas de Gran Canaria*, 10–14. Retrieved August 29, 2011 from <http://www.lrec-conf.org/proceedings/lrec2002/pdf/ws8.pdf>

Software and electronic tools

- Anthony, Laurence. 2010. AntConc (Version 3.2.0m) [Computer software]. Tokyo: Laurence Anthony. Retrieved June 7, 2010 from http://www.antlab.sci.waseda.ac.jp/antconc_index.html
- Scott, Mike. 1996. WordSmith Tools [Computer software]. Oxford: Oxford University Press.
- Scott, Mike. 2010. WordSmith Tools (Version 5.0) [Computer software]. Oxford: Oxford University Press. Retrieved June 5, 2010 from <http://www.lexically.net/wordsmith/index.html>
- Wufoo.com. Palo Alto: SurveyMonkey.com, LLC.