



Jan Kochanowski University Press

This is a contribution from *Token: A Journal of English Linguistics*
Volume 4/2015.

Edited by John G. Newman, Marina Dossena and Sylwester Łodej.

© 2015 Jan Kochanowski University Press.

Parallel and comparable corpora in investigating modal verbs in legal and literary discourse

Bogumiła Pacyna

Adam Mickiewicz University, Poznań

ABSTRACT

This article offers a qualitative and quantitative analysis of modal verbs that were found in an online corpus, called the Lagun Corpus. The case of modal verbs has been the subject of much debate in the literature not limited only to one type of discourse. In this paper various texts are analysed; however, the primary focus is on legal and literary texts. The analysis developed in this paper shows the dominance of one modal and explains the rejection of others. It also gives an account of the usefulness of corpora in translation research and seeks to determine the most appropriate corpora for translators. In particular, the paper highlights parallel and comparable corpora as those most relevant to translators.

Keywords: modal verbs, corpora in translation studies

1. Introduction

This article aims to add to the discussion on the nature of modality in the English language. It outlines the results of a case study conducted on English modal verbs. The case study is divided into three parts: the first describes the data, the corpus used to collect the data, the aim of this case study and discusses its research questions. Parts two and three show the results of the case study and answer the research questions. Moreover, the theoretical background for the paper presents the usefulness of corpora in translation research.

2. Corpora in translation

Kübler – Aston (2010: 501-502) state that corpora can be of use in the three stages of a translation process, that is, documentation, drafting, and revision. First, in the documentation process, corpora provide a set of texts from a range of various disciplines, and are thus a source of specialised terminology and constituents in a given discourse. Second, in the drafting process as explained by Varantola, corpora may give “reassurance when checking hunches or finding equivalents translators are not familiar with” (2000: 118). Finally, in the revision process corpora can confirm “readability, comprehensibility, coherence, grammaticality or terminological consistency” (Kübler – Aston 2010: 501-502). The observations presented by Kübler – Aston also hold for corpora in translator training. It is emphasised that corpora may

develop awareness of cross-cultural similarities and differences, of strategic alternatives and of the role of context. They can also improve linguistic and world knowledge, help in the acquisition of new concepts or new uses and develop awareness of the technical issues involved in computer-assisted translation (Kübler – Aston 2010: 512).

Kenning (2010: 494) mentions that translation scholars utilise corpora in translation research, first of all, to estimate the impact a source text may have on a target text, the so called translationese phenomenon. Some forms and structures may be adopted in the target text, such as discourse conventions or complexity. Moreover, scholars would also describe elements that distinguish a translated language from a non-translated one – see Baker (1996: 176-177).

Following Kenning (2010: 492-493), it is worth noting that corpora can be applied in contrastive linguistics to analyse lexical items, syntax, or discourse. Scholars claim that corpora can be used to describe languages more precisely, to show the degree to which some patterns are shared by languages as well as serving a good function in testing a pre-formed hypothesis.

Kenning (2010: 487) enumerates two types of corpora that are particularly useful to translation scholars: parallel and comparable corpora. A parallel corpus can be utilised in translation training, bilingual lexicography and machine translation (Kenny 2001: 51). Furthermore, such a corpus may “provide information on language-pair specific translational behavior, to posit certain equivalence relationships between lexical items or structures

in source and target languages (Kenny 1998; Marinai – Peters – Picchi 1992) or to study the phenomenon of translationese” (Schmied – Schäffler 1996). Comparable corpora are believed to be most useful to translation scholars as they provide accurate information about translated texts (Kenny 2001: 53). Kübler – Aston (2010: 510) suggest that it may be difficult to create “a strictly comparable corpora for specialized domains”. There may not be enough texts in languages different from English. The question that arises here is *how comparable and analysable* the corpora can be. Kenning draws the conclusion that all the merits and drawbacks confirm scholars in the belief that balancing between corpora is the best alternative, emphasising that “exploiting several corpora leads to a fuller and more accurate account of the phenomenon under investigation” (2010: 496-497).

3. Case study

3.1 Data and the aim of the research

The data for this study were collected from an online parallel corpus, called the Lagun Corpus, comprising Basque, Spanish, French, Polish and English. Only the English-Polish part was used in the study where English was the source text. The study was conducted in 2012. At that moment, it was an experimental version of the tool. The corpus functioned with a simplified search engine; however, despite its narrow remit, it was still possible to obtain interesting results from it. The corpus included texts from literature, such as prose, theatre, religion, (1,178,138 words), journalism (14,126 words), scientific, such as human science, exact science, social science and natural sciences, (408,402 words) and technical texts, such as juridical texts, (436,575 words): overall 2,037,241 words.

The purpose of this corpus study is to test the feasibility of the following research questions: (1) Which modal verbs are most frequently used in the English part of the corpus? (2) In which of the particular English texts are modal verbs most frequent? (3) Which meaning of *shall* and *would* is most frequent? (4) How are *shall* and *would* translated from English into Polish in the Lagun Corpus?

Three groups of modal verbs were tested in this corpus study: central, marginal and semi-modals, following the division proposed by Depraeter – Reed (2006: 269). The first group comprises *can*, *could*, *may*, *might*, *shall*,

should, will, would and *must*. The second group comprises *dare, need* and *ought to*. The third group comprises *have to, be able to, be going to, be supposed to, be about to* and *be bound to*.

3.2 The frequency of modal verbs

Table 1 shows the findings of the corpus study concerning the frequency of use of the modal verbs. The table reveals the instances of modal verbs found in scientific texts, technical texts, literature, and journalism. The total body of research consists of 11,390 instances of verbs. *Must* and *must not* were analysed separately and are presented in separate columns, bearing in mind the essential difference in meaning.

Table 1. Modal verbs found in four types of discourse

Modal verb	Records found	Literature	Scientific texts	Technical texts	Journalism
can	815	599	176	40	–
could	710	667	39	2	2
may	716	81	101	534	–
might	226	190	22	14	–
shall	2059	96	32	1931	–
should	410	268	111	31	–
will	572	359	163	38	–
would	1023	925	81	12	5
must	425	293	54	78	–
must not	15	6	4	5	–
dare	44	37	7	–	–
need	229	130	51	41	7
ought to	71	54	17	–	–
have to	152	130	19	2	1
be able to	96	78	11	5	–
be going to	112	112	–	–	–
be supposed to	15	13	2	–	–
be about to	34	34	–	–	–
be bound to	8	2	6	–	–
Total	11390	7734	896	2745	15

Firstly, the study revealed that the most frequent modal verbs are *shall*, with over 2000 records, and *would*, with over 1000 records. *Can*, *could* and *may* are also used very often, with over 700 records found for these verbs. Approximately 500 records were found for *should*, *will* and *must*. Less frequent verbs were *might*, *need*, *have to* and *be going to* where 226, 229, 152 and 112 records were found respectively. The least common verbs are *dare*, *ought to*, *be able to*, *be supposed to*, *be about to* and *be bound to*. Fewer than 100 records were found for these verbs.

Secondly, literature appears to be the most common type of discourse in which modal verbs are used. However, it should be taken into account that over 57% of the total words collected in the corpus come from literature. Likewise, technical texts and scientific texts are abundant in modal verbs. The smallest number of instances were found in journalistic texts. This is due to the low number of data from journalism.

Moreover, three central modals appear frequent in literature, namely *would*, *can* and *could*. *May* and *shall* are used in technical texts more often than in literature or scientific texts. Verbs which were found in fewer than 500 instances will not be discussed here. Another conclusion is that semi-modals are rare in use in all forms of texts.

Furthermore, it may be observed that *shall* is one of the most frequently used verbs and it is mostly found in juridical texts (technical texts). In these texts *shall* is used to establish rights and duties.

Finally, it is remarkable that only a few instances of modal verbs were found in journalistic discourse. Modal verbs in newspaper articles comprise less than one per cent of the total number of records. This anomaly can be accounted for by the small number of data from journalism (only 14,126 words).

3.3 What is *shall* and *would* most common meaning?

The main objective of this part of the corpus study is to reveal the most common use of *shall* and *would* in literature and juridical texts. Moreover, translations of *shall* and *would* from the Lagun Corpus will be examined. Table 2 presents the main uses of *shall* and *would* and their frequency in literature and juridical texts.

The data suggest that *shall* is more common in juridical texts and its use is restricted to formal instructions. Over 1800 instances with this meaning were found. In only about 80 instances was *shall* used to express prohibition. On the other hand, *shall* in literature is much rarer and is used to express certainty, suggestions and intentions.

Table 2. *Shall* and *would* in literature and juridical texts

Modal verb and its use		Literature	Juridical texts (Technical texts)
SHALL	1) certainty	50	–
	2) prohibition	–	79
	3) formal instructions	–	1852
	4) suggestions	22	–
	5) intentions	24	–
WOULD	1) regularity in the past	251	–
	2) thinking about the future	383	12
	3) reported speech	179	–
	4) unwillingness	1	–
	5) assumption	2	–
	6) instructions and appeals	2	–
	7) offers and invitations	11	–
	8) regret	–	–
	9) politeness	1	–
	10) question tags	4	–
	11) wishes	63	–
	12) preference	5	–
	13) certainty	21	–

As can be seen from the table above, *would* is mainly used to express future, regularity in the past or in reported speech. Moreover, it is apparent from this table that *would* occurs in very few instances with the meaning of certainty, wish or offer. A small number of records representing unwillingness, assumption, instructions, politeness or question tags were noted. No records were found for regret. Finally, as Table 2 shows, there is a significant difference between literature and legal discourse, where only twelve records of *would* were identified.

3.4 Translation

This section provides translations of sentences from the Lagun Corpus. The examples present English sentences with *shall* and *would* as well as their Polish translations. The first two examples come from legal documents. The remaining six examples come from literature.

- (1) Everyone **shall contribute** to sustain public expenditure according to their economic capacity, through a fair tax system based on the principles of equality and progressive taxation, which in no case shall be of a confiscatory scope.
Wszyscy **wnoszą** wkład w podnoszenie podatków publicznych zgodnie ze swoimi możliwościami za pośrednictwem sprawiedliwego systemu podatkowego inspirowanego przez zasady równości i progresywności, który w żadnym wypadku nie może przybrać rozmiarów konfiskaty.
- (2) The sum of the amounts deemed necessary, fixed in the specific programmes **shall not exceed** the overall maximum amount fixed for the framework programme and each activity.
Suma kwot uznanych za niezbędne, ustalonych przez programy szczegółowe, **nie może przekroczyć** łącznej maksymalnej kwoty przewidzianej dla programu ramowego i dla każdego działania.
- (3) My mother sometimes tells me that never in my whole life **shall** I ever again **see** rivers as beautiful and big and wild as these, the Mekong and its tributaries going down to the sea, the great regions of water soon to disappear into the caves of ocean.
Matka mówi mi czasem, że nigdy w życiu **nie zobaczę** rzek równie pięknych jak właśnie ta, równie wielkich i dzikich jak Mekong i jego odnogi spływające do oceanu, wodne obszary, które nikną w głębinach mórz.
- (4) “**Shall** we **listen** to the news?” Sarah asked.
Posłuchamy wiadomości? Zapytała Sara.
- (5) She said, “I **shall do** what he tells me to do”.
Zrobię to, co mi każe.
- (6) I was frightened but I knew I **would** never **give up**.
Bałem się, ale wiedziałem, że się nigdy **nie poddam**.
- (7) It happened every day. Of that I’m sure. It must have come on quite suddenly. At a given moment every day the despair **would make its appearance**.
Tak musiało być co dzień, mam co do tego pewność. Musiało to być okrutne. Codziennie w pewnym momencie **zjawiała się** ta rozpacz.

- (8) The woman went in ahead of me and immediately started to remove the newspapers very carefully from the furniture. I told her I **would do** that myself. She ignored me. I raised my voice and told her to quit it, I wanted to be alone to relax, I had had a long journey.

Kobieta weszła przodem i natychmiast z wielką uwagą zaczęła ściągać gazety z mebli. Powiedziałem jej, że sam to zrobię. Nie słuchała. Powtórzyłem głośno żeby zostawiła, że po długiej podróży chcę zostać sam, w spokoju.

These few examples show that English modals can be translated differently into Polish. In most translations there are no Polish modals used, e.g. in sentences (1), (4), (5), (7) and (8). A modal verb appears only in (2). However, the translations transmit a similar spirit to the original sentences. Moreover, the translations are easily comprehensible and read well as if they were originally written in the target language. There are no spelling errors or missing words. That proves the accuracy of the translations.

4. Conclusions

Returning to the questions posed at the beginning of this study, it is now possible to state that *shall* and *would* are the most common modal verbs used in the English part of the Lagun Corpus. It has also been noted that the frequency of occurrence of *dare*, *ought to*, *be able to*, *be supposed to*, *be about to*, and *be bound to* is low and therefore these modals are not analysed in this study.

Moreover, it has been shown that *shall* is mainly used to express formal instruction and *would* carries future meaning. These modal verbs appear most often in two types of texts, namely literary and juridical texts. *Shall* is more common in juridical texts whereas *would* in literary texts.

The final observations pertain to the translation of modals from English into Polish in the Lagun Corpus. It has been shown that in most of the translations modal verbs do not occur but still the sentences convey the same feeling to the source sentences. These findings enhance our understanding of modal verbs and may serve as a base for the future study of the subject.

This paper has also given an account of and the reason for the widespread use of corpora in translation studies. In short, corpora can be useful for the translation process and provide a whole gamut of functional vocabulary items or confirm initial considerations.

Additionally, contrastive linguists may satisfactorily compare languages thanks to corpora. It has been proved that parallel and comparable corpora appear to be most helpful tools for translators. However, it has been shown that complimentary use of the corpora is of a real benefit.

REFERENCES

Sources

Lagun Corpus

2010 <http://korpus.hiztegia.org/> (last available in 2011)

Special studies

Baker, Mona

1996 "Corpus-based translation studies: The challenges that lie ahead". In: Harold Somers (ed.) *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam: Benjamins, 175-186.

Depraetere, Isle – Susan Reed

2006 "Mood and modality in English". In: Bas Aarts – April McMahon (eds.) *The Handbook of English Linguistics*. Oxford: Blackwell, 269-290.

Kenning, Marie-Madeleine

2010 "What are parallel and comparable corpora and how can we use them?" In: Anne O'Keeffe – Michael McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 487- 500.

Kenny, Dorothy

1998 "Theme and rheme in Irish and English: A corpus-based study". In: *School of Applied Language and Intercultural Studies Working Papers in Language and Society 1998*. Dublin: Dublin City University, 1-25.

2001 "Corpora in translation studies". In: Mona Baker (ed.) *Routledge Encyclopedia of Translation Studies*. London: Routledge, 50- 53.

Kübler, Natalie – Guy Aston

2010 "Using corpora in translation". In: Anne O'Keeffe – Michael McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 501-515.

Marinai Elisabetta – Carol Peters – Eugenio Picchi

1992 "Bilingual reference corpora: creation, querying, applications". In: Ferencz Kiefer – Gabor Kiss – Julia Pajzs (eds.) *Papers in*

Computational Lexicography Complex 92. Budapest: Linguistic Institute, Hungarian Academy of Sciences.

Schmied, Josef – Hildegard Schäffler

- 1996 "Approaching translationese through parallel and translation corpora". In: Ian Lancashire – Carol Percy – Charles Meyer (eds.) *Studies in Synchronic Corpus Linguistics*. Amsterdam: Rodopi, 41-55.

Varantola, Krista

- 2000 "Translators, dictionaries and text corpora". In: Silvia Bernardini – Federico Zanettin (eds.) *I Corpora nella didattica della traduzione/ Corpus Use and Learning to Translate*. Bologna: CLUEB.

Address: BOGUMIŁA PACYNA, Faculty of English, Adam Mickiewicz University in Poznań, al. Niepodległości 4, 61-874 Poznań, Poland.