Journal of Professional and Scientific Communication

Vol. XLIV 3–4/2022

FACHSPRACHE

facultas

Herausgeber/Editors: Prof. Dr. Jan Engberg (je@cc.au.dk),
Prof. Dr. Ines-Andrea Busch-Lauer (Ines.Busch.Lauer@fh-zwickau.de),
Prof. Dr. Nina Janich (nina.janich@tu-darmstadt.de), Prof. Dr. Merja Koskela (merja.koskela@uwasa.fi)
Redaktion & Manuskripte/Editorial Secretary & Submit Papers:
Dr. Anja Steinhauer (fachsprache@facultas.at), Guidelines: https://ejournals.facultas.at/index.php/fachsprache

Eigentümer und Verleger/Proprietor and Publisher:
Facultas Verlags- und Buchhandels AG, facultas, 1050 Wien, Österreich
Vorstand/Managing Director: Dr. Rüdiger Salat
facultas.at

Erscheinungsweise/Publication Details: zwei Doppelhefte pro Jahr mit Beiträgen in Deutsch, Englisch, Französisch und Spanisch/two double issues a year, contributions in German, English, French and Spanish
Preise/Prices: Jahresabo PREMIUM (print & online, inkl. Archiv & E-only-Sonderheft)/
annual subscription PREMIUM (print & online, incl. access to archives & e-only special edition) EUR 103,–
Jahresabo für Studierende PREMIUM (print & online, inkl. Archiv & E-only-Sonderheft)/
annual subscription PREMIUM for students (print & online, incl. access archives & e-only special edition) EUR 69,–
Jahresabo/annual subscription CLASSIC (print) EUR 72,–
Jahresabo für Studierende/annual subscription for students CLASSIC (print) EUR 47,–
Einzelheft (print oder pdf) sowie Sonderheft (pdf)/single issue (print or pdf) or special issue (pdf) EUR 38,–
Beitrag/article (pdf) EUR 24,–
Institutslizenz/institutional membership EUR 264,–
(Preise inkl. USt., zzgl. Versand/prices incl. taxes, plus shipping costs)
Bestellung/Orders: ejournals@facultas.at

Kontakt/Contact: Facultas Verlags- und Buchhandels AG, facultas,
Stolberggasse 26, 1050 Wien, Österreich, Tel.: 0043 1 310 53 56, Fax: 0043 1 319 70 50,
facultas.at bzw. https://ejournals.facultas.at/index.php/fachsprache
Anzeigen/Advertisements, Website & Newsletter: Katharina Schindl, ejournals@facultas.at

Layout & Satz: Beate Soltész, www.soltesz-grafik.at
Druck: Facultas AG, 1050 Wien, Österreich

ISSN: 1017-3285

Liebe Leserinnen und Leser,

wieder einmal gibt es etwas Neues bei FACHSPRACHE. In Herausgeberteam und Verlagsredaktion diskutieren wir jedes Jahr gemeinschaftlich, welche unserer Rubriken welchen möglichen Bedarf in der *scientific community* abdecken und wie sich die zunehmende Digitalisierung unserer Lebenswelt und der Wissenschaft auf solche Bedarfe auswirkt. In diesem Zuge haben wir schon vor längerer Zeit die Rezensionen aus dem Printheft ausgekoppelt und in Form einer eigenen Rubrik („Reviews") auf die Homepage verschoben, kürzlich dann auch die von Ines Busch-Lauer sorgsam erstellte fortlaufende Bibliografie.

Letztere ist ein mühsames Unterfangen – zudem wird sie zunehmend von automatisierten Verfahren der Verlinkung und Aktualisierung von Forschung im Internet überholt. Weiterhin halten wir aber Rezensionen als Instrument der Information und Orientierung für wichtig, auch wenn es immer schwieriger wird, Kolleg:innen für diesen Dienst an der Gemeinschaft gewinnen zu können. Die vor einigen Jahren geschaffene Kategorie „Disputationen", in der Nachwuchswissenschaftler:innen ihre kurz vor der Publikation stehenden Dissertationen vorstellen können, möchten wir aus ähnlichen Gründen beibehalten.

Dennoch wollen wir der durch das Internet forcierten Dynamik wissenschaftlicher Angebote noch besser gerecht werden. Wir nehmen daher im Herbst 2022 folgende Änderungen am Online-Angebot vor:

- Statt der Rubrik „Reviews" gibt es jetzt eine mit dem Namen „Forschung aktuell / Research Update". Hier finden Sie zukünftig – chronologisch geordnet und kontinuierlich aktualisiert – Kurztext-Angebote zu neuen Entwicklungen in der Forschung: Rezensionen, Disputationen, Tagungsberichte sowie Webinare mit Diskussionen zu aktuellen Veröffentlichungen bzw. Themenheften.
- Die Fortführung der Bibliografie wird eingestellt, die Bibliografie bleibt aber als historisches Dokument weiterhin auf der Homepage von Fachsprache erhalten und einsehbar (Stand 2021). Ebenso bleibt das Archiv der Rezensionen erhalten.

Damit entlasten wir die Printversion von FACHSPRACHE noch weiter zugunsten der wissenschaftlichen Originalpublikationen und ermöglichen es Ihnen zugleich, auch außerhalb des halbjährlichen Heft-Erscheinungstermins auf Informationen rund um aktuelle Forschung zugreifen zu können. Die in FACHSPRACHE publizierten Artikel stehen ihnen schon jetzt nach jeweils 12 Monaten auf der Homepage kostenlos im Volltext zur Verfügung (https://ejournals.facultas.at/index.php/fachsprache).

Wir danken Ihnen, wenn Sie uns auch weiter als Leser:innen, Beiträger:innen, Reviewer:innen und Abonnent:innen treu bleiben – Ihr Herausgeberteam

Dear Readers,

FACHSPRACHE again announces something NEW to make your reading experience even better! The editorial board and the publishing house discuss every year about which journal sections best meet the ever-changing requirements of our *scientific community* and how the enormous digital transformation that is happening in our lives and in science will impact such needs. As a result of our discussions, we decided, already some time ago, to change the publication of reviews from print to online in a section entitled ("Reviews"), and more recently to also make the ongoing bibliography of LSP research, carefully compiled by Ines Busch-Lauer, available via the publisher's homepage.

Bibliography compilation is labour-intensive – what's more it is increasingly being overtaken by automated procedures for linking and updating research on the Internet. However, we still consider reviews as important as a tool for information and orientation, even if it is becoming increasingly difficult to attract colleagues for review services to the community. For similar reasons, we would like to retain the section "Disputations", which was created a few years ago and allows young researchers to present their dissertations that are about to be published.

We are striving to even better meet the dynamic requirements of the Internet when offering research findings. Starting in autumn 2022, the following changes will be made in our online offering:

- "Reviews" will be replaced by a section entitled "Forschung aktuell / Research Update". In the future, this section will include – chronologically ordered and continuously updated – all short text offerings on new development strands in research: reviews, disputations, conference reports as well as webinars with discussions on current publications or special issues.
- The bibliography will be discontinued, but its previous issues will be preserved as archived document (as of 2021) and available on the FACHSPRACHE homepage. Likewise, the review archive will remain online.

In this way, we are improving the print version of FACHSPRACHE in favour of original scientific publications, while at the same time enabling you to access information about current research outside of the biannual issue publication date. The articles published in FACHSPRACHE are already made available free of charge as full text on the homepage after every 12 months, cf. https://ejournals.facultas.at/index.php/fachsprache.

We thank you for remaining loyal to us as readers, contributors, reviewers, and subscribers.

Sincerely,

Your editorial team

The Features of Legal Language in British Legislation: a Diachronic Perspective

Eva Dvořáková

Abstract The aim of this paper is to explore diachronically the features of legal language in British legislation over a period of 200 years to find out whether the features identified by Crystal and Davy in their seminal work are still present in current British legislation and whether any changes can be attributed to plain language efforts. Further, the paper explores lexical diversity and readability. For these purposes, five corpora of British legislation were compiled from 1820, 1870, 1920, 1970, and 2020. The results show that the current language of British legislation is very different from the legal language described by Crystal and Davy fifty years ago. The plainer texts tend to be lexically less diverse. The readability measures provide rather inconclusive results.

Keywords archaisms, British legislation, legal language, lexical diversity, plain language, readability

1 Introduction

When asked to identify typical features of legal language, most people will mention long, complex sentences, accurate terminology, and certain legalisms, such as *shall*, *hereinafter*, or *aforesaid*. However, do these perceptions match the reality?

In their seminal work, *Investigating English Style*, in 1969, Crystal and Davy offered a perspicacious insight into legal language, summarising its main features by analysing two extracts from legal documents (Crystal/Davy 1990). However, when looking at the two extracts, one wonders whether such examples are representative of legal language today, fifty years after the book was published. In other words, plain language movements[1] and other factors may have changed legal language to the point where the description provided by Crystal and Davy may now only partially apply.

The aim of this paper is to examine diachronically a genre[2] of legal English, namely British legislation, to find out whether it possesses the features of legal language described by Crystal and Davy and whether these features have changed over time. The selected features are lexical features that can be analysed through corpora. They are the following: the scarcity of personal pronouns, the use of archaic compound adverbs, pairs of synonymous adjectives, the modal

[1] Some authors disagree with the word *movement*. For example, according to Balmford (2002), plain language "has grown beyond being a movement to become a product, a business, an industry, or a professional service".

[2] This paper uses the word *genre* as understood by Kurzon: "We recognize texts belonging to specific legal genres – contracts, judgments, legislation – through pragmatic means, by way of looking at the purpose of the text." (Kurzon 1997: 125)

Zitiervorschlag / Citation:
Dvořáková, Eva (2022): „The Features of Legal Language in British Legislation: a Diachronic Perspective." *Fachsprache. Journal of Professional and Scientific Communication* 44.3–4: 112–129.

shall, and the word *such* used as a determiner. These features will be analysed through corpora of British legislation from 1820, 1870, 1920, 1970, and 2020.

Over the past decades, the move towards plain language has influenced the way official and legal documents are drafted. For example, Wydick (1978: 738) recommends using familiar words and avoiding lawyerisms. In this way, by restricting the vocabulary, legal language could become lexically less diverse than language that is not subject to such restrictions. Recent technological developments have made it possible to assess lexical diversity relatively easily. The relevant diachronic corpora will be analysed to support or disprove the hypothesis that the introduction of plain language tends to make texts lexically less diverse.

Finally, the avowed aims of plain language exponents are to make official and legal documents more accessible to their readers. The development of so-called readability measures has facilitated the objective assessment of texts in terms of their accessibility to lay readers. Most of the measures take account of sentence length and the number of syllables in individual words; some check the vocabulary against a list of frequently used words. Thus, the third strand of this research aims to analyse our diachronic corpora of British legislation by means of readability measures, to determine whether these objective tools can confirm that over time British legislation has become plainer and therefore more accessible to lay readers.

In 1969, Crystal and Davy selected two extracts from legal documents (an endowment assurance policy and a hire purchase agreement) as the basis for their analysis. Those were private documents from the domain of contract law. Our study is limited to British legislation. Thus, the distinction between the drafting practices in public and private documents may have affected the results. Nevertheless, Crystal and Davy apparently used their two texts to derive general principles applicable to legal language as such, as evidenced by their explanation that they "have chosen two examples that … [they] felt to be reasonably central in a linguistic sense" (Crystal/Davy 1990: 195). This raises the question of whether the texts selected for the purposes of the present study, namely legislation, can be considered sufficiently "central" so as to be representative of legal language as a whole. We believe that legislation is one of the central genres which sets the pace for the whole domain of legal language. Yet the existence of multiple genres within legal language makes any generalisation problematic; the paper therefore does not aspire to cover the whole of legal language, but rather to explore solely British legislation.

While insightful research into several aspects of plain language in legislation has been undertaken over the last decades (for example the use of *shall* was explored by Garzone [2013a, 2013b] and Williams [2006, 2012]), to the best of our knowledge little effort has been made so far to diachronically compare British legislation over a substantial period of time, using corpora and data. We therefore hope to contribute to a better understanding of the practical benefits of plain language.

2 Literature review

In the 1960s Mellinkoff offered, in his seminal work *The Language of the Law*, a comprehensive analysis of legal language. Rather than applauding the peculiarities of this specific type of language, he vigorously criticised certain typical features which tend to hamper comprehension or are simply redundant. With his statement "the language of the law should not be different without a reason" (Mellinkoff 1963: 285), he embraced a pragmatic approach that readers should not be required to struggle in order to understand legal documents. A similar spirit can be found in the books by Butt (2018) and Tiersma (2000).

In 1969, Crystal and Davy offered their dispassionate analysis of legal language in the book *Investigating English Style*. Rather than taking a stance like Mellinkoff or later Butt and Tiersma, they sought to objectively identify the features of legal language, by analysing two extracts from legal documents (an endowment assurance policy and a hire purchase agreement) that they selected as representative of legal language. The authors provided a comprehensive analysis of legal language, starting with the layout and capitalisation, continuing with the vocabulary, and ending with the sentence structure.

Other influential scholars and professionals (e. g., Garner 2001, Kimble 1992, Williams 2015) facilitated the spread of these ideas across the globe and in professional settings. In this way, the plain language requirement has found its way into statutes (e. g., The US Plain Writing Act of 2010) and numerous legislative drafting manuals.[3] Yet "regulatory demand" is not the only driving force; in Australia, plain language tends to be introduced in response to "client demand" (Balmford 2002: 5.4), being seen as a competitive edge for some law firms: "One day, clients everywhere will refuse to pay for legal services unless they are plain" (Balmford 2002: 5.3). This is in stark contrast to the "bang for the buck" argument (Sneddon 2011: 713) explained by Gopen: "Clients who pay such prices, the argument runs, want to see their received value in terms of the degree of difficulty of the product." (Gopen 1987: 345)

The adoption of plain language statutes requires some guidance for judges to determine which texts are plain and compliant. Cheek (2010) offered a comprehensive approach to plain language definitions, comprising the following: (1) numerical or formula-based definitions (using the Flesch readability formula); (2) elements-focused definitions; and (3) outcomes-focused definitions. The formula-based definitions provide only "rough guides, however, scores derived from readability formulas provide quick, easy help" (DuBay 2004: 19). Cheek prefers outcomes-focused definitions which can, however, be rather demanding in terms of their implementation. She further suggested that "plain English is not an absolute but should be appropriate to the intended audience" (Cheek 2010: 9). In fact, consumers cannot be seen as the only addressees of plain language because "clear communication is for all" (Balmford 2002: 5.1), and even judges seem to prefer plain language (Palyga 1999). Nevertheless, Long and Christensen (2011) have shown that briefs written in plain language do not have an impact on the outcome of the case. Conversely, Benson and Kessler have established that "[l]awyers who write in legalese are likely to have their work judged as unpersuasive and substantively weak" (Benson/Kessler 1987: 319).

The advice given by plain language promoters to lawyers is to use familiar words and avoid lawyerisms so as not to "send your reader groping for the dictionary" (Wydick 1978: 738). Can such an approach have an impact on the lexical diversity of legal texts? Cvrček and Chlumská (2015) examined one of the translation universals, namely simplification, in translated texts, trying to find an answer to the question of whether translated texts tend to be lexically less diverse. To do that, the authors sought to develop a reliable method for measuring lexical diversity, given the limitations of the traditional TTR (type-token ratio) approach. They came up with the zTTR method (a modified type-token ratio approach, incorporated into the online tool https://www.korpus.cz/calc/) which compares the TTR values of the examined text with referential values, with the text type and text size also being taken into consideration. This novel method subsequently confirmed their hypothesis that the translation process tends to deprive translated texts of some of their lexical diversity. Although the present study does not

[3] For more details on the most important plain language initiatives, cf. Williams 2015.

deal with translations, the methodology used by Cvrček and Chlumská could help to determine whether plain language texts tend to be lexically poorer.

Since the 1940s, economic interests have led book and newspaper publishers to try to increase their readership by specifically tailoring the texts to the skills of the readers; this led to the development of objective measures of readability, the so-called readability formulas. In his comprehensive overview of readability principles, DuBay (2004) provided valuable insights into numerous readability formulas. The best-known formula is the Flesch formula, based on sentence length and the number of syllables. A different approach was adopted for the Dale-Chall formula which, apart from sentence length, uses a list of 3,000 easy words and detects the "hard" words not included in the list. DuBay mentions a study according to which "the average adult in the U.S. reads at the 7th-grade level" (DuBay 2004: 1) and materials for public should be written at fifth- or sixth-grade reading level. The research into readability thus provides relatively objective criteria for the drafters of official documents who seek to adapt their messages to their readership. Yet as suggested by DuBay, readability measures have their limitations. This has been aptly demonstrated by Šlerka and Smolík (2010) who applied various readability formulas to a variety of Czech texts. Among other things, they showed that a children's book scored as the most difficult text from their corpus.

3 Methodology

3.1 Corpora

When designing the research, the obvious question arises as to how long the time span to be investigated should be. Since the key results should elucidate how the language of modern British legislation has changed in comparison with the legal language of the time when *Investigating English Style* by Crystal and Davy was published (1969), the key time periods are 1970 and 2020. In this way, it is possible to see how legal language has developed over the past 50 years and whether the plain language campaigns have had any measurable impact on the language of British legislation. In addition, however, it would be interesting to see the bigger picture and find out whether some of the changes to legal language would have occurred anyway. Consequently, it was decided to cover a longer time span, namely 200 years.

The first part of the research required the compilation of corpora of British legislation. It was decided to create 5 corpora from the following years: 1820, 1870, 1920, 1970, and 2020. The selected statutes were public general acts, rather than regulations or statutory instruments. Delegated legislation was excluded because the drafting style might conform to different drafting rules than primary legislation. Primary legislation is drafted by the Office of Parliamentary Counsel, whereas "delegated legislation is instructed and drafted within each Government Department" (Xanthaki 2013a: 58). Another criterion was that no amendments were to be included. The drafters of amendments tend to adopt the same language as in the original statutes, so the amendments are not necessarily representative of the specific period of time. The next criterion was purely technical – the format of the statutes was PDF, not scanned documents which would not be analysable through corpus managers. This applied primarily to the oldest legislation. Furthermore, in order to be included in the corpus, a statute had to contain at least 3 paragraphs of text – again, this seemed to be particularly relevant in the earliest stages when the statutes were often rather short. The last criterion was that statutes were used in their original "as enacted" form, not including any amendments from later peri-

ods. The question then arose as to how large the corpora had to be and whether they should be of comparable sizes. Given the substantial differences in length – new legislation tends to be much longer than old legislation – it was decided to use 100 statutes in total, 20 from each period. The whole corpus contains 974,780 tokens. The earliest sub-corpora account for about 10 % each, while the largest subcorpus represents 40 % of the whole corpus size. This composition is presumably more likely to yield representative results than a corpus which would contain only very few statutes from 2020. Overall, the results are presented in instances per million (ipm) which ensures that the results are comparable across corpora of different sizes.

3.2 Research questions

The overall aim of this paper is to show how British legislation has changed over the last 200 years with respect to certain legal language features identified by Crystal and Davy (1990), whether the trend towards plainer English tends to deprive legal texts of their lexical diversity, and whether the readability scores can confirm that British legislation has become more accessible to ordinary citizens. The corpus analysis was conducted via Sketch Engine (https://www.sketchengine.eu/).

Crystal and Davy presented a comprehensive description of legal language, discussing the layout, vocabulary, and sentence structure. In this paper we decided to focus only on vocabulary for several reasons. Firstly, a study of this size cannot offer a detailed comparison of all the features identified by Crystal and Davy (1990). Secondly, certain features may be inherently linked to the specific genre that Crystal and Davy analysed; e. g., the layout of an endowment assurance policy may be incomparable with the layout of an act of Parliament. Thirdly, the efforts of plain language exponents have largely targeted the lexis of legal documents; consequently, a diachronic comparison of the lexical changes can explain whether these efforts have yielded any tangible benefits. Fourthly, our methodology is corpus-based, and not all the features mentioned by Crystal and Davy are analysable via corpora.

Among other things, Crystal and Davy have noted the "extreme scarcity" (Crystal/Davy 1990: 202) of pronouns and anaphora, resulting in substantial repetitions of lexical items. Firstly, the analysis should show whether there is a noticeable trend regarding the use of personal pronouns over the years. Secondly, the paper seeks to find out whether British legislation has lost some of its archaic features. Crystal and Davy provided three examples of archaisms: the ending *-eth* (as in *witnesseth*), the compound adverbs of the type *hereinbefore*, and the word *aforesaid* (Crystal/Davy 1990: 206 f.). The combination of a corpus-based and corpus-driven approach concentrated on these three types of archaisms and their development over time. Thirdly, coordinated synonymous adjectives were explored, using a computer query to automatically generate coordinated items which were subsequently manually checked to exclude the items which are not synonymous (e. g., *public and private*). The last part of the analysis examines the frequencies of *shall* and *such*. In the case of *such*, the query has to be formulated so as to yield the results for use as a determiner ("unaccompanied by the indefinite article" [Crystal/Davy 1990: 206]). In Sketch Engine, the tag NN covers singular or mass nouns. Ideally, the query should not cover mass nouns (as in *such information*), but only the cases where *such* is combined immediately with a countable singular noun. Nevertheless, this drawback should not substantially distort the results, because the main focus is on the relative frequencies over time, and if mass nouns are included in all time periods, we should still be able to see the trend.

The second part of the research aims to analyse the texts in terms of their readability. For that purpose, we selected 2 readability formulas (Flesch Reading Ease Test and Dale-Chall formula) that can be used on the website https://readabilityformulas.com/. The Flesch Reading Ease Test concentrates on sentence length and word length, while the Dale-Chall formula compares sentence length and the "hard" words which are not included in the list of 3,000 "easy" words. As the website allows texts of a maximum length of 3,000 words, we used 5 random samples from each corpus.

The third part of the research explores lexical diversity. The Czech National Corpus has developed an online tool called the corpus calculator (https://www.korpus.cz/calc/) which makes it possible to determine the lexical diversity of a text, based on the type/token ratio, taking into consideration the length and type of a text. After entering the values for the number of types and the number of tokens, the tool creates a bell curve with two vertical lines: the expected value (in the middle) and the observed value. The position of the observed value indicates whether the text is lexically more or less diverse than expected for the specific type and length of text. The aim is to find out whether the most recent corpus is lexically less diverse than the oldest corpus.

4 Results

4.1 Personal pronouns

With regard to personal pronouns, Figure 1 confirms the "scarcity" of personal pronouns. If general English corpora are used as reference corpora (BNC and English Web 2020), it becomes obvious that personal pronouns are used substantially more in general English than in the legislation corpus. Over time, there is a noticeable downward trend, with the 1820 corpus containing more than twice as many personal pronouns per million (ipm = instances per million) as the 2020 corpus.

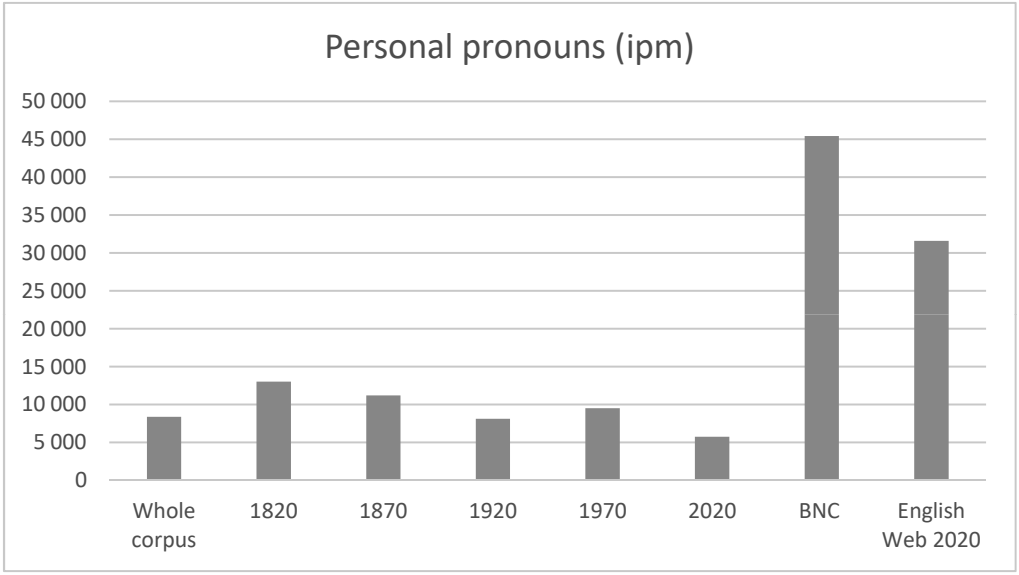

Figure 1: Use of personal pronouns in legislation and in general English

Overall, this paper confirms the finding made by Crystal and Davy (1990) that personal pronouns tend to be eschewed as a species in legal language. Furthermore, from the diachronic perspective we have noted a marked trend towards the use of fewer personal pronouns in legislation over time.

4.2 Archaisms

The first group of archaisms (the ending *-eth)* yielded no results. The ending *-th* was slightly more prevalent, covering the word forms *hath* and *doth*, all of them being from 1820 (ipm 19).

The second group of archaisms (Figure 2), derived from *here-/there-/where* + preposition, is relatively large, composed of 43 compound words. In the literature, these words are referred to as textual mapping adverbials (Gotti 2012: 56), compound adverbs or deictic pro-forms (Quirk et al. 1985: 438, 487), some of them as conjunctive adverbs (Halliday/Hasan 1976: 230), or simply as *here-* and *there-* words (Butt 2018: 286, 638). The "here" part of the word has a pronominal function (for example, *thereof* means 'of that document') rather than a locative meaning; in fact, it denotes only proximity or distance. Interestingly, although it has been established that personal pronouns "seem to be eschewed as a species" in legal texts (Crystal/Davy 1990: 202), the *here-/there-/where-* words (with a pronominal function) have flourished as characteristic features of legal texts for a long time. Over the last few decades, they have been fiercely targeted by the plain language exponents (Tiersma 2000: 94, Butt 2013: 236, Garner 2001: 401).

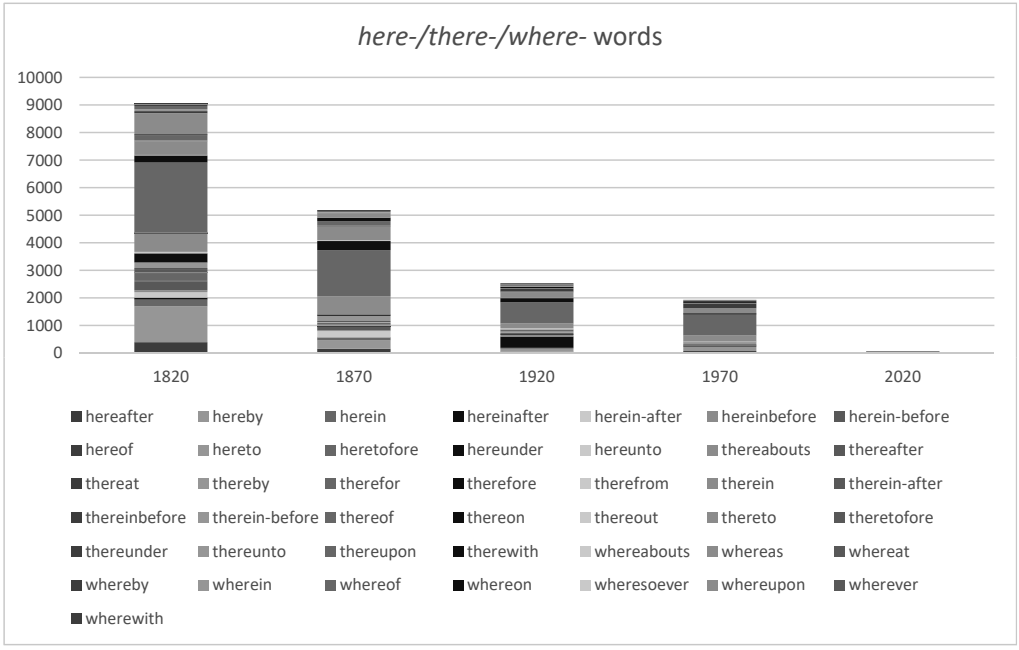

Figure 2: Use of here-/there-/where-*words in legislation from 1820 to 2020*

The distribution of the words across all 5 corpora is uneven; while the 1820 corpus totals over 9,000 *here-/there-/where*-words (ipm), the 2020 corpus contains only 54 (ipm) of them. The

chart indicates that every 50 years the numbers of archaic words halved, with the exception of the corpora from 1920 and 1970 where the changes were less distinct, and the 2020 corpus where the numbers plummeted from 1,929 *here-/there-/where-* words to 54 (ipm). Certain *here-/there-/where-* words were extremely frequent in the earliest corpora, and thus have seen the greatest declines. For example, the words *hereby, therein, thereof,* and *whereas* all decreased from very high numbers in 1820 to almost zero in 2020.

Around 1969, when Crystal and Davy's book was published, the *here-/there-/where-* words still had a relatively high frequency in our corpora. It is therefore hardly surprising that this feature was identified by those authors as characteristic of legal language. Yet our data show that even back then the drop from the earliest corpus (1820) had been very steep, from around 9000 ipm to 2000 ipm (in 1970), suggesting that the archaic *here-/there-/where-* words were already doomed at the time when *Investigating English Style* was published. Every fifty years the ipm of the *here-/there-/where-* words substantially dropped. The 2020 corpus is virtually free of compound adverbs and these words can thus no longer be regarded as a typical feature of British legislation. Does it mean that once these archaic words have been removed, the text is easier for non-experts? Masson and Waldron (1994) have undertaken a study to examine which changes make a legal text (a contract) more accessible to non-experts. They had four versions of a legal document: (1) the original, (2) with archaisms removed, (3) with plainer language – shorter sentences, simpler syntax, etc. and (4) with special terminology replaced by easier words or explained. After checking the comprehension by the subjects, the authors found that solely the removal of archaisms did not significantly improve comprehension. It seems, therefore, that the mere removal of archaisms in legislation does not guarantee greater comprehension by non-experts.

The prefix *afore-* was used only in the word *aforesaid* (Figure 3) which peaked in 1870 (over 3,000 ipm) and then steadily decreased to 0 in 2020.

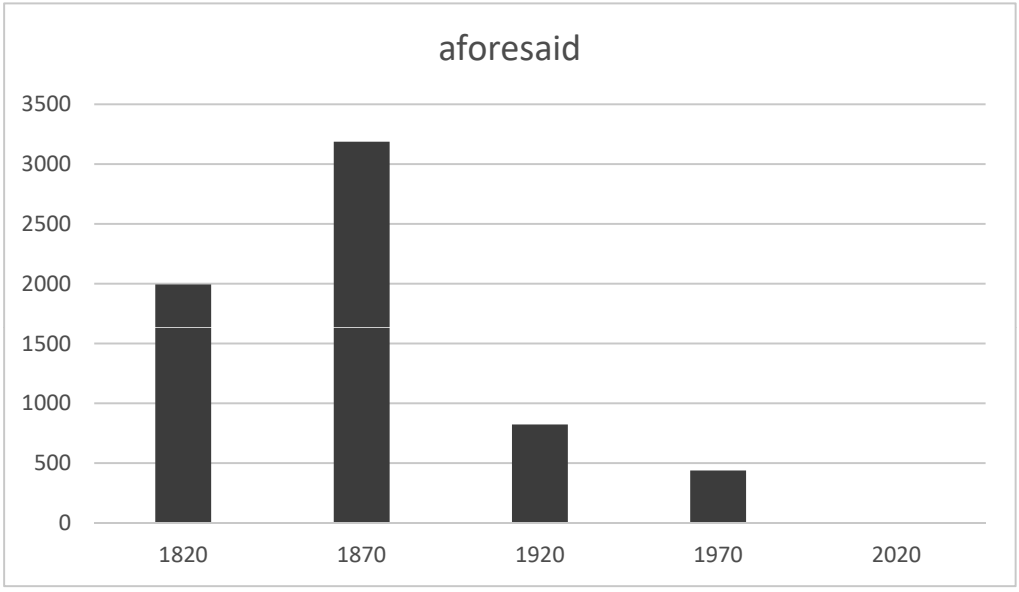

Figure 3: Use of aforesaid *in legislation from 1820 to 2020*

The word *aforesaid* was used as an adjective with the meaning 'stated before'. According to Tiersma (2000: 90), this word had been introduced into legal English from Latin (*predictus*) or Law French (*le dit*) and can be easily replaced by *the* or *this*. A typical collocation from our 1820 corpus is *the aforesaid act* (7 occurrences). The 2020 corpus contains no instances of *aforesaid* and the comparable collocations with *act* contain the words *that* or *the*.

4.3 Coordinated adjectives (binomials)

Figure 4 lists the synonymous coordinated adjectives from our corpora. The 1820 corpus contains as many as 16 distinct pairs, and every successive time period has seen a substantial reduction. The 2020 corpus has only one pair.

Coordinated adjectives	**1820**	separate and distinct	**1870**	due and proper
		valid and effectual		valid and effectual
		null and void		good and effectual
		due and payable		absolute and unfettered
		final and conclusive		usual and ordinary
		due and owing		valid and binding
		good and perfect		due and payable
		requisite and necessary		final and conclusive
		full and complete		
		just and equitable		
		further and additional		
		answerable and responsible		
		good and lawful		
		good and effectual		
		willing and desirous		
		due and unpaid		
	1920	efficient and economical	**1970**	due and unpaid
		final and conclusive		due and outstanding
		fair and equitable		final and conclusive
		suitable and convenient		
		valid and effectual	**2020**	efficient and effective
		frivolous and vexatious		

Figure 4: Use of coordinated adjectives in legislation from 1820 to 2020

In terms of meaning, these expressions have largely become lexicalised and can be treated as a single idiomatic unit. Although the binomials are traditionally not included under standard word-formation processes, they serve to enrich legal vocabulary through the "lexicalisation of syntactic phrases" (Kastovsky 2006: 209). The motivation behind their use was varied: they

could cover synonymous words of different origin (Anglo-Saxon, Latin, French) and thus facilitate communication in a multilingual society, they could serve to achieve precision by specifying the vague first item through the second item (Dobrić Basaneže 2018: 204), they could have had different meanings in the past which later converged, as in *null and void* (Mellinkoff 1963: 358), or they could serve to emphasise the ritual and ceremonial aspects of legal instruments (Mellinkoff 1963: 92). However, as Wydick argued, they are a "lawyer's tautology – a needless string of words with the same or nearly the same meaning" (Wydick 1978: 734) and as such they are discouraged in modern legal language. Semantically, these expressions often denoted *validity* and *debt*: e. g., *final and conclusive*, and *due and payable*.

4.4 Shall

The iconic[4] word *shall* (Figure 5) is a clear example of a very frequent word in legislation which became virtually extinct in the 2020 corpus. With regard to the dynamic situation around *shall*, Williams speaks of a "modal revolution" (Williams 2012: 363) and many experts on legislative drafting recommend that this "chameleon-hued word" (Garner 2001: 939) be abandoned altogether.[5] Others make the case for "a disciplined use of *shall*" (Adams 2014: 12). For example, according to the "American rule" (Garner 2001: 940), *shall* should only be used where it means "has a duty to". Other scholars maintain that "the replacement of *shall* with other forms […] does not always bring a real improvement" (Garzone 2013a: 115). Interestingly, Mellinkoff, the forefather of the plain language movement, did not specifically target *shall*. He even used it in his improved re-drafted version of a lease provision (Mellinkoff 1963: 388).

The word *shall* was particularly targeted by the plain language exponents because of its ambiguity. It can impose an obligation, grant a right, impose a condition precedent/subsequent, state a fact or assumption, and denote futurity (Garner 2001: 940, Butt 2018: 565). Others see the "degree of duty" as problematic (Kimble 1992: 61). According to Garzone, polysemy is a "distinctive property of all modals" (Garzone 2013a: 99) and thus every substitute modal for *shall* has the potential for ambiguity as well. Furthermore, Garzone challenged the view that *shall* should be used only when an obligation is imposed on a person. She argued that in addition to the deontic meaning, *shall* has a performative/constitutive value (e. g., "The declaration *shall apply to …*") (Garzone 2013a: 99), thus casting doubt on the "has a duty" test. Kurzon brought the perspective of speech acts, exploring the difference between the phrases "The Director *shall* give to the Committee …" and "The Director *has the duty* to give to the Committee …". Kurzon contends that the former case is a speech act serving as a command, while the latter ("has the duty") "is the result of an order having been given. It is in fact a description of a state of affairs" (Kurzon 1986: 22). In other words, from the perspective of speech acts, the uses of *shall* and *has a duty* are not identical.

As can be seen in Figure 5, the current British statutes are practically free of *shall*, although 50 years earlier, the ipm was still very high (almost 10,000). If we consider the entire 200-year period, the steady downward trend is apparent throughout (with a minor deviation in the 1920 corpus where the numbers are slightly higher than in the 1870 corpus), and the greatest decrease occurred between 1970 and 2020.

---

4 According to Kimble, "*shall* is the most important word in the world of legal drafting […] *shall* is the most misused word in the legal vocabulary" (Kimble 1992: 61).

5 "*Shall* has had its day" (Butt 2018: 565); "My suggestion is to abandon *shall* altogether" (Asprey 1992: 79).

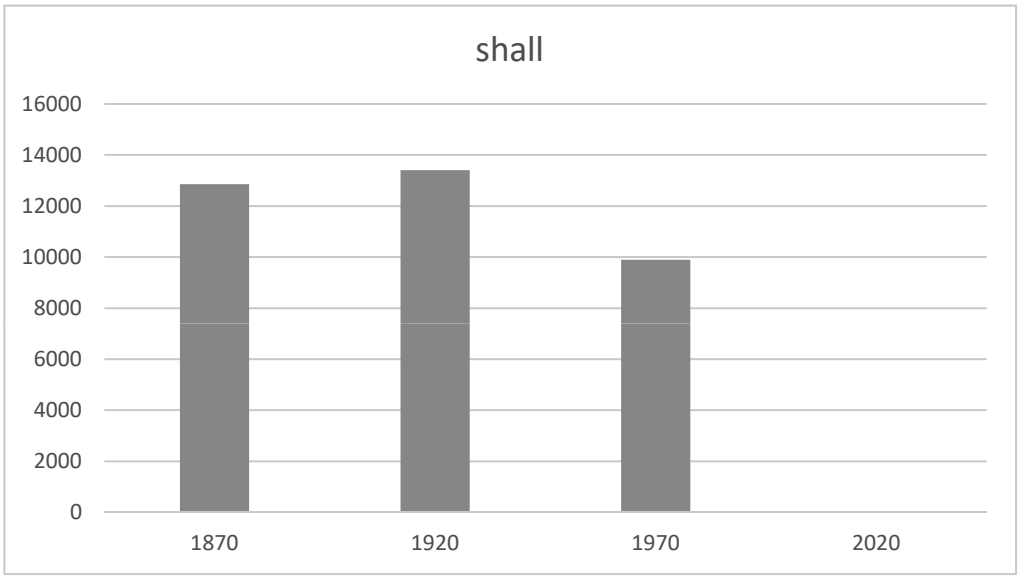

Figure 5: Use of shall *in legislation from 1820 to 2020*

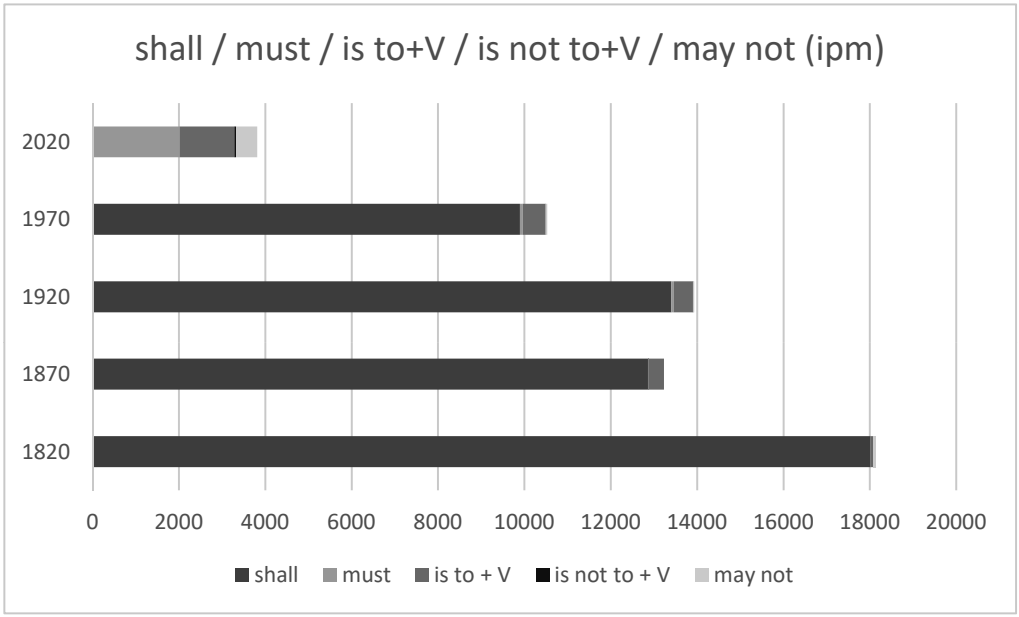

Figure 6: Use of alternatives to shall *in legislation from 1820 to 2020*

Such an enormous reduction in *shall* raises the question of whether it has been substituted with some other words, or whether it was truly redundant and could be easily removed without any substitution. The *Drafting Techniques Group Paper 19* (2008) provided a list of different uses of *shall* and the substitution strategies in various contexts. The authors suggested the following alternatives: *must* for imposing obligations, *there is to be* for the creation of statutory

bodies, and the present tense in provisions about application or effect, amendments, repeals, and other common provisions. Williams proposed the substitutes according to the "hierarchy of normative intensity" (Williams 2012: 366): *must* for maximum strength, followed by *is to / are to* and the present tense. In our corpora, we examined the use of the most straightforward alternatives to *shall*: *must, is to, may not*. It would be expected that fewer occurrences of *shall* mean substantially more occurrences of these alternatives. As Figure 6 shows, this is not the case.

While there is a considerable increase in *must, is to* + verb, and *may not* in 2020, the rows are by no means comparable in terms of their length. Thus, it is necessary to look at other substitution strategies. In many civil law countries legislation is drafted in the present tense. According to Xanthaki, "legislation does not need to repeat that its text is compulsory: irrespective of the use of an imperative form, legislation is inherently compulsory" (Xanthaki 2013b: 104). Consequently, with the decreasing use of *shall* over time, we expect to find an increased use of the present tense. Arguably the best way to determine the use of the present tense is to focus on the 3rd person singular (with an *-s*). The corpus analysis yielded the following results.

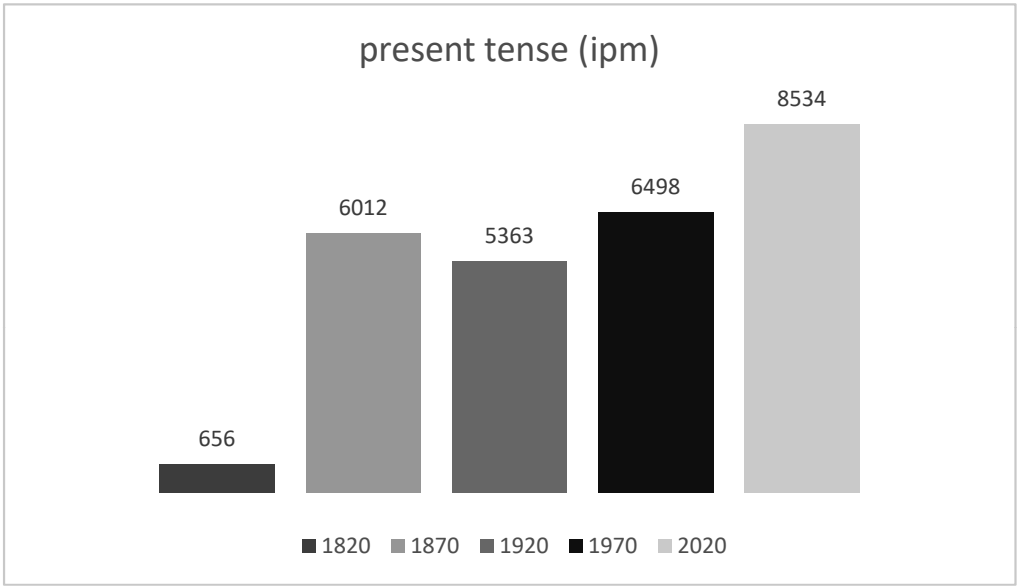

Figure 7: Use of the present tense (3rd person singular) in legislation from 1820 to 2020

As can be seen in Figure 7, in the 1820 corpus the present tense (the third person singular) was very rare, and within 50 years it had become ten times more frequent. In fact, the slight deviation in 1920 (a decreased use of the present tense) corresponds to the deviation in Figure 6 where the 1920 corpus shows an increased use of *shall*. Similar findings over a 50-year period were identified by Garzone: "In theoretical terms, the simple present is the best substitute for *shall*" (Garzone 2013a: 109). Yet Garzone (2013b: 77) gives an example of a provision in which the replacement of *shall* with the present tense changes the meaning: "There *shall* be a body corporate …" is different from "There is a body corporate …" in that the latter case implies that the body is already in existence. Overall, Garzone (2013b: 79) argues that the replacement of

shall with other forms poses difficulties because the substitutes can have a slightly different meaning from *shall* and can be fuzzy and ambiguous as well.

4.5 Such

Figure 8 shows the use of the word *such* as a determiner (without an indefinite article). It peaked in the 1870 corpus, and is probably heading towards extinction in British legislation. Although the values in the 2020 corpus are still significant (498 ipm), compared to the initial values in the 1820 corpus (4814 ipm), this represents a tenfold drop.

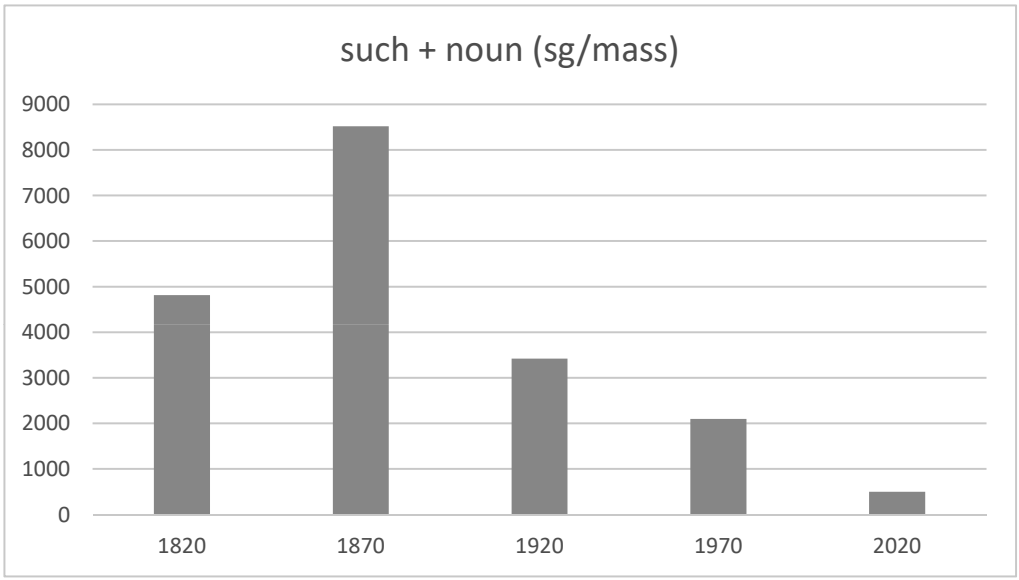

Figure 8: Use of such *without an indefinite article from 1820 to 2020*

The substantial decline shows that the "precision of reference" which was suggested by Crystal and Davy as a possible reason for using *such* was a myth, and precision could be easily achieved by other means, for example the definite article.

4.6 Readability

The readability scores for the Dale-Chall formula and the Flesch Reading Ease formula (5 samples for each corpus) were rather inconclusive and failed to prove that the current British legislation is more readable than the old statutes. Overall, the readers of British legislation need tertiary education (Grades 13–15, 16+) to understand the texts (18 samples out of 25) which indicates that the texts cannot be classified as readable. The clearest texts seem to be those from the oldest corpora.

The two readability formulas thus failed to confirm the hypothesis that the newer legislative texts, which are arguably plainer in terms of their lexis, are more readable. Several factors could account for this:

(1) Contemporary society deals with more complex and technical issues than in the 1820s, and the law needs to reflect this. In this way, while certain "glue words" (Wydick 1978: 729) have been removed from legislation, the "working words" and terminology may have become more sophisticated and technical. Likewise, the complexity and interrelatedness of legal concepts may require more complex sentence structures.
(2) While in recent decades legislative drafters have focused on the vocabulary of the legal texts to make it more compliant with plain language principles, sentence structure may have been slightly neglected. Thus, it could be argued that the mere substitution of individual words does not necessarily ensure greater readability. For example, Garzone maintains that the replacement of *shall* "is an easily accomplished change within the context of a much more complex and long overdue process of simplification of legal texts, while more profound improvements at the level of syntax and text organization are certainly more difficult to achieve" (Garzone 2013b: 79). In addition, there could be some other factors that present difficulties for the layperson. Azuelos-Atias (2018) mentioned three problematic areas: technical vocabulary, syntactic complexity, and intertextuality (implied background knowledge) which requires "detective work" (Azuelos-Atias 2018: 106) on the part of the lay person.
(3) Readability formulas have their limitations – generally they take into account the length of sentences and the length of words, but a long sentence (if properly structured) need not necessarily pose substantial difficulties for the reader.

From the synchronic perspective, all British legislation as a whole seems to be accessible primarily to a university-educated readership. So, if an average adult reads at the 7th-grade level (DuBay 2004: 1), British legislation is largely inaccessible to him/her. This begs the question of addressees: whom are the statutes drafted for? Even experts do not seem clear on this issue, and over time opinions have changed to a certain degree. For example, 70 years ago Driedger believed that the addressees of legislation were professionals and experts, not ordinary citizens: "It must not be supposed, however, that statutes can be written so that everyone can understand them" (Driedger 1949: 295). The more recent approach seems to take the view that if everyone is presumed to know the law, then the law should be written in a form accessible to everyone. Thus, many promoters of plain legal language (e. g., Tiersma, Butt, Garner) subscribe to Mellinkoff's view that "the language of the law should not be different without a reason" (Mellinkoff 1963: 285). But the fundamental question is whether legislation *can* indeed be written for the average adult who reads at a 7th-grade level. After all, a legislative text must meet many criteria that other types of text do not have to fulfil.

4.7 Lexical diversity

Figure 9 shows the lexical diversity of the oldest (1820) and the most recent (2020) corpus, with the expected value (the vertical line in the middle) and the observed value (the vertical line on the left). Both corpora are lexically less diverse than expected for this type of text (as evidenced by the shorter line which appears on the left rather than on the right). From the diachronic perspective, the most recent corpus (2020) is lexically less diverse than the 1820 corpus, with the observed value being about one third of the expected value.

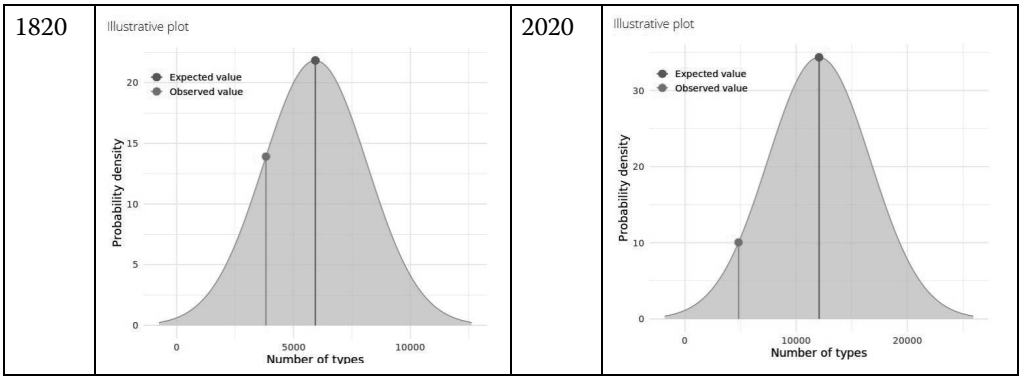

Figure 9: Lexical diversity of the 1820 and the 2020 corpora

The *Drafting Guidance* of the Office of the Parliamentary Counsel mentions the advice given by Sir Ernest Gowers[6] on the choice of words in legislation: "use the most familiar words" (Office of the Parliamentary Counsel 2020: 5). By limiting the choice of words to the most familiar words and by consciously avoiding traditional and archaic words, the lexical diversity of a text can be affected. This was confirmed in this paper, as the most recent corpus (2020) is clearly lexically less diverse than the older corpus. Nevertheless, it would require further research to find out whether plain language generally results in a lexically less diverse text, as there could be some other factors at play.

5 Impact of plain language

Most of the substantial changes that have moulded the language of British legislation into its present form seem to be part of a long trend stretching over the past 200 years. Thus, already at the time when Crystal and Davy were writing their seminal book, some of the features identified by them as typical of legal language were already doomed and on their way to extinction (at least as far as British legislation is concerned), probably due to consistent language planning.[7] However, the greatest change for many of the analysed features was made between 1970 and 2020. The use of the *here-/there-/where-* words fell steadily from 1820 onwards, but the presence of such words in legislation was nevertheless substantial until 1970 (1,929 ipm). Only in the 2020 corpus can we see that this group of words was almost completely eradicated. The same applies to *aforesaid*, whose presence was steadily declining, but some presence of *aforesaid*, albeit minor, was still expected even in 2020. Yet the 2020 corpus does not contain a single instance of *aforesaid*. The word *shall* also confirms that considerable acceleration[8] took place after 1970; in the 1970 corpus, the presence of *shall* was high (9,898 ipm), while in the 2020 corpus there were only 23 instances per million. This is probably no coincidence; since

6 The author of *Plain Words: A Guide to the Use of English*, published in 1948, well before the plain language movements took the form of a concerted effort.

7 See Williams: "Prescriptive engineering", "the changes may have occurred as a result of nurture rather than nature" (Williams 2012: 356).

8 Garzone and Williams do not seem to agree on the moment when the greatest reductions in *shall* occurred; while Williams (2012) places the greatest reductions between 2001 and 2010, Garzone suggests that "the decline may have started before 2001" (Garzone 2013b: 70).

the 1970s plain language exponents have globally exerted a considerable influence on legal language, seeking to remove certain problematic features which could impede comprehension by non-lawyers.

The language of legislation can be substantially influenced from the top, by means of drafting guidelines. In the UK, plain language rules have found their way into the *Drafting Guidance* of the Office of the Parliamentary Counsel. In addition to explicitly referring to certain promoters of plain language (e. g., Asprey), the guide advises on how to use certain traditional legal words. For example, the version from June 2020 expressly recommends avoiding *shall* and archaisms (Office of the Parliamentary Counsel 2020: 4 f.). Since British primary legislation is produced centrally via the Office of Parliamentary Counsel, the *Drafting Guidance* certainly plays an important role in ensuring a high degree of standardisation. In fact, it appears to be a very effective tool for introducing changes to the language of legislation, and thus, compared to other genres of legal language, the language of legislation seems to be rather dynamic.

Another important factor is the entire system of legislative drafting in common law countries where legislation tends to be drafted in one central drafting unit (Stefanou 2016). In the case of the United Kingdom, it is the Office of Parliamentary Counsel. Logically, if all the legislative drafts originate from a single drafting unit, a high degree of standardisation can be relatively easily achieved, and if there is consensus about the introduction of plain language elements into legislation, this change can be quite rapidly implemented in this centralised environment.

6 Conclusion

This paper sought to explore certain lexical features of legal language, as presented by Crystal and Davy in their seminal work *Investigating English Style,* in British legislation from five different periods (1820, 1870, 1920, 1970, and 2020). Our analysis showed that many features identified by Crystal and Davy as typical of legal language (e. g., archaisms, *shall*, the word *such* used as a determiner, pairs of synonymous adjectives) are virtually non-existent in current British legislation. The data from the 2020 corpus show that external factors must be behind the accelerated disappearance of some of the features. These external factors are most likely the plain language efforts which have exerted considerable influence on the language of legislation via the *Drafting Guidance* of the Office of the Parliamentary Counsel.

Because plain language promoters often tend to recommend restricting the choice of vocabulary to the most frequently used words, this paper also explored the lexical diversity of the oldest and the most recent corpus to find out whether the development of legal language has brought about lexically less diverse texts. This hypothesis was confirmed, as the 2020 corpus tends to be lexically less diverse than the 1820 corpus.

The last part of our research concentrated on the readability of British legislation to find out whether the changes made to the language of legislation have made British statutes more accessible to non-lawyers, as far as the readability measures can show. The readability scores from the five samples from each corpus were rather inconclusive (in fact, some scores cautiously suggested that the older texts were more readable), thus the only tentative conclusion that can be drawn in this respect is that (1) the current legislative texts may deal with more technical matters requiring more specialised (and less readable) language; (2) the recent plain language efforts may have focused on vocabulary, neglecting sentence structure and other issues; and (3) readability formulas have their limitations and cannot provide reliable results.

Further research is needed to objectively measure the accessibility of legislation to non-lawyers.

It follows from our analysis that the recent decades have seen dramatic changes in the language of British legislation, which may have affected perceptions about the typical features of legal language. These changes are largely due to language planning efforts. It remains to be seen whether such changes will gradually trickle down to private legal documents as well. Likewise, it will be interesting to see whether changes to the register will gradually affect the perceptions of law students and the general public who often still see legal language as the one presented by Crystal and Davy fifty years ago.

References

Adams, Kenneth A. (2014): "Banishing Shall from Business Contracts: Throwing the Baby out with the Bathwater." *Australian Corporate Lawyer* 24: 12 f.

Asprey, Michele M. (1992): "Shall must go." *The Scribes Journal of Legal Writing* 3: 79–83.

Azuelos-Atias, Sol (2018): "Making Legal Language Clear to Legal Laypersons." *Legal Pragmatics*. Eds. Barbara Kryk-Kastovsky and Dennis Kurzon. Amsterdam: Benjamins. 101–115.

Balmford, Christopher (2002): "Plain Language: Beyond a Movement: Repositioning Clear Communication in the Minds of Decision-makers." *Plainlanguage.gov website:* https://www.plainlanguage.gov/resources/articles/beyond-a-movement/ (31.08.2022).

Benson, Robert W. / Kessler, Joan B. (1987): "Legalese v. Plain English: An Empirical Study of Persuasion and Credibility in Appellate Brief Writing." *Loyola of Los Angeles Law Review* 20.2: 301–322.

Butt, Peter (2013): *Modern Legal Drafting: A Guide to Using Clearer Language.* Cambridge University Press.

Butt, Peter (2018): *Legal Usage. A Modern Style Guide*. LexisNexis Butterworths.

Cheek, Annetta (2010): "Defining Plain Language." *Clarity* 64: 5–15.

Crystal, David / Davy, Derek (1969): *Investigating English style*. London: Routledge.

Crystal, David / Davy, Derek (1990): "The Language of Legal Documents." *Investigating English Style*. Eds. David Crystal / Derek Davy. 12th impres. Harlow: Longman. 193–217.

Cvrček, Václav / Chlumská, Lucie (2015): "Simplification in Translated Czech: A New Approach to Type-token Ratio." *Russian Linguistics* 39.3: 309–325.

Dobrić Basaneže, Katja (2018): "Extended Binomial Expressions in the Language of Contracts." *Phraseology in Legal and Institutional Settings*. Eds. Anne Wagner / Vijay Kumar Bhatia / Stanisław Goźdź-Roszkowski / Gianluca Pontrandolfo. London: Routledge. 203–220.

Drafting Techniques Group (2008): *Shall*. Paper 19 (final).

Driedger, Elmer A. (1949): "Legislative Drafting." *Canadian Bar Review* 27: 291–317.

DuBay, William H. (2004): *The principles of Readability*. ERIC Clearinghouse. https://files.eric.ed.gov/fulltext/ED490073.pdf (04.10.2022).

Garner, Bryan A. (2001): *A Dictionary of Modern Legal Usage*. 2nd ed. New York: Oxford University Press.

Garzone, Giuliana (2013a): "Modality and Performativity in Legislative Texts: The Case of Shall." *The Three Waves of Globalization: Winds of Change in Professional, Institutional and Academic Genres*. Eds. Franca Poppi / Winnie Cheng. Newcastle upon Tyne: Cambridge Scholars Publishing. 94–122.

Garzone, Giuliana (2013b): "Variation in the Use of Modality in Legislative Texts: Focus on Shall." *Journal of Pragmatics* 57: 68–81.

Gopen, George D. (1987): "The State of Legal Writing: Res Ipsa Loquitur." *Michigan Law Review* 86.2: 333–380.

Gotti, Maurizio (2012): "Text and Genre." The Oxford Handbook of Language and Law. Eds. Peter Meijes Tiersma / Lawrence M. Solan. Oxford / New York: Oxford University Press. 52–66.

Halliday, Michael Alexander Kirkwood / Hasan, Ruqaiya (1976): *Cohesion in English*. London: Longman.

Kastovsky, Dieter (2006): "Vocabulary." *A History of the English Language*. Eds. Richard Hogg / David Denison. Cambridge: Cambridge University Press. 199–270.

Kimble, Joseph (1992): "The many Misuses of Shall." *The Scribes Journal of Legal Writing* 3: 61–77.

Kurzon, Dennis (1986): *It is hereby performed … Explorations in Legal Speech Acts.* Amsterdam: Benjamins.

Kurzon, Dennis (1997): "'Legal Language': Varieties, Genres, Registers, Discourses." *International Journal of Applied Linguistics* 7.2: 119–139.

Long, Lance N. / Christensen, William F. (2011): "Does the Readability of Your Brief Affect Your Chance of Winning an Appeal? An Analysis of Readability in Appellate Briefs and its Correlation with Success on Appeal." *The Journal of Appellate Practice and Process* 12.1: 145–162.

Masson, Michael E. J. / Waldron, Mary Anne (1994): "Comprehension of Legal Contracts by Non-experts: Effectiveness of Plain Language Redrafting." *Applied Cognitive Psychology* 8.1: 67–85.

Mellinkoff, David (1963): *The Language of the Law*. Eugene: Wipf and Stock. (June 2020). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892409/OPC_drafting_guidance_June_2020-1.pdf (03.09.2022).

Palyga, Steve (1999): "Is it Safer to Use Legalese or Plain English? What the Judges Say." *Clarity* 43: 46–51.

Phillips, Alfred (2003): *Lawyers' Language: How and why Legal Language is Different*. London: Routledge.

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

Šlerka, Josef / Smolík, Filip (2010): "Automatická měřítka čitelnosti pro česky psané texty." *Studie z aplikované lingvistiky / Studies in Applied Linguistics* 1: 33–44.

Sneddon, Karen (2011): "Speaking for the Dead: Voice in Last Wills and Testaments." *St. John's Law Review* 85.2: 683–754.

Stefanou, Constantin (2016): "Comparative Legislative Drafting: Comparing across Legal Systems." *European Journal of Law Reform* 18.2: 123–138.

Tiersma, Peter M. (2000): Legal Language. Chicago: The University of Chicago Press.

Williams, Christopher (2006): "Fuzziness in Legal English: What Shall we Do with Shall?" *Legal Language and the Search for Clarity*. Eds. Anne Wagner / Sophie Cacciaguidi-Fahy. Bern: Lang. 237–264.

Williams, Christopher (2012): "Changes in the Verb Phrase in Legislative Language in English." *The Verb Phrase in English: Investigating Recent Language Change with Corpora*. Ed. Bas Aarts. Cambridge: Cambridge University Press. 353–371.

Williams, Christopher (2015): "Changing with the Times: The Evolution of Plain Language in the Legal Sphere." *Revista Alicantina de Estudios Ingleses* 28: 183–203.

Wydick, Richard C. (1978): "Plain English for Lawyers." *California Law Review* 66.4: 727–765.

Xanthaki, Helen (2013a): "Legislative Drafting: A New Sub-Discipline of Law Is Born." *IALS Student Law Review* 1.1: 57–70.

Xanthaki, Helen (2013b): *Thornton's Legislative Drafting*. London: Bloomsbury.

Eva Dvořáková
Faculty of Law, Charles University in Prague
nám. Curieových 7, 116 40 Staré Město,
Faculty of Arts, Charles University in Prague
nám. J. Palacha 1/2, 116 38 Staré Město,
e-mail: dvorakoe@prf.cuni.cz

Maritime Cooperative Working Agreements: Variability as a Proxy for Legal Atomization

Mary C. Lavissière & Laurent Fedi

Abstract Textual standardization is the expected route for documents in specialized disciplinary fields, particularly for legal documents. In this paper, however, we question this representation, contradicting literature that represents contracts as "frozen" genres. We argue that variation in certain legal documents is a result of complex legal contexts. We carry out a textometric study of 46 Cooperative Working Agreements in the maritime industry. We observe substantial differences in the lexical content of these contracts, despite the fact that they are all considered technical contracts. We argue that this linguistic variability stems from the lack of a uniform legal framework for the maritime industry. We conclude that the fluid legal framework and the variety of agreements that result from it are exploited to meet the parties' objectives under the guise of purely technical collaboration.

Keywords contracts, discourse analysis, English for Specific Purposes, genre analysis, legal English

1 Introduction

In language for specific purposes, textual standardization is the expected route for documents in specific disciplinary fields. This is especially true of legal genres because compliance with textual norms is strongly characteristic of legal genres in general, and, specifically, of contracts (Gotti 2012). Indeed, standardization underlies the very notion of genre. A genre can be defined as an event with a recognizable and accepted – i. e. standardized – form that allows for legitimately reaching a professional goal (Orts Llopis 2014). This is especially true of contracts. We argue, therefore, that a simplified representation of contractual language as fixed and highly-standardized is erroneous. Instead, we defend the view that variability and innovation are also crucial to the functionality of professional language (Bhatia 2012), including contractual language. Additionally, our study adds depth to this second vision of contracts by bringing a description of Cooperative Working Agreements (CWAs), which are currently understudied from a genre perspective. Second, we show that the variability in contractual language stems from a legal framework that is atomized. This variability in language allows for private companies to meet their latent objectives under the guise of purely technical collaboration.

These cooperation tools are classified as technical contracts with an operational dimension (Ghorbani et al. 2021) and not commercial contracts (Fedi/Tourneur 2015, Brooks et al. 2019, Fedi et al. 2022). Yet they have a substantial impact on the maritime transport market (Merk/Teodoro 2022) since they constitute transboundary agreements around the world. They have been identified as highly controversial from legal, economic, and political perspectives

Zitiervorschlag / Citation:
Lavissière, Mary C. / Fedi, Laurent (2022): „Maritime Cooperative Working Agreements: Variability as a Proxy for Legal Atomization." *Fachsprache. Journal of Professional and Scientific Communication* 44.3–4: 130–147.

(ITF 2018, 2019). Most recently, they came under sharp criticism during President Joe Biden's 2022 *State of the Union Address*,

> We see it happening with ocean carriers moving goods in and out of America. During the pandemic, these foreign-owned companies raised prices by as much as 1,000 % and made record profits. Tonight, I'm announcing a crackdown on these companies overcharging American businesses and consumers. (The White House 2022)

The American President's words were followed by a press release from the American Department of Justice and the Federal Maritime Commission (FMC), signaling increased scrutiny about the way these companies work under the framework outlined in CWAs (Justice Department 2022). These remarks are the latest of a series of criticisms about the impact of maritime CWAs from international, American, and Chinese business and legal authorities. The recent civil penalty of 2 million USD against the fifth largest container liner company, clearly shows the abuses of the sector regarding demurrage and detention fees (FMC 2022b).

From a business perspective, CWAs aim to rationalize risks and costs in maritime shipping, which has been characterized as a volatile, risky, and competitive sector (Stopford 2009, Caschili et al. 2014). Notwithstanding the significant work of harmonization undertaken by the International Maritime Organization (IMO) with regard to safety, security and environmental protection (IMO 2020), the legal context remains atomized. Neither domestic nor international law can govern all disputes (Orts Llopis 2014). In response to this context, actors in the industry have begun working together in legal arrangements governed by CWAs (Caschili et al. 2014). As private documents, however, the contents of most CWAs have remained opaque. This opacity continues today even though American law (Shipping Act 1984, Ocean Shipping Reform Act 1998) mandates that liner companies operating in American jurisdictions file their agreements with the FMC. The FMC then makes these documents public through their website. Despite this availability, however, no linguistic studies have yet to be published on the nature of CWAs and their various forms. This is a significant gap because these instruments underpin global shipping alliances and main trade routes of containers in the world (Fedi et al. 2022).

Our study aims to fill this gap. It is based on a corpus of 46 CWAs from the container liner shipping sector. The agreements were filed with the FMC by international actors in the industry between 1985 and 2020. They are published on the FMC's website (FMC 2022a). Using textometry through the software program Iramuteq, we analyze the variation in lexical content according to their secondary classification as determined by the FMC and to their effective date as determined by this authority.

Our study is organized as follows. First, we include a brief overview of the literature about genre and variation in legal language and, specifically, in contracts. We also survey the literature on the historical and legal contexts of CWAs. Second, we describe our corpus of CWAs. We further explain the textometric methods we used for the study. Third, we present the results of the textometric analysis. Fourth, we discuss how our results contribute to the current literature. Fifth, we conclude with the limits of our study and perspectives for future research.

To facilitate reading, we provide a table of the abbreviations used in the article.

Table 1: Abbreviations used in the article

Full form	**Abbreviation**
Cooperative Working Agreement	CWA
Digital Container Shipping Association Agreement	DCSAA
Descending Hierarchical Classification	DHC
Federal Maritime Commission	FMC
Research Question	RQ
Slot Charter Agreement	SCA
Vessel Sharing Agreement	VSA

2 Literature review and research questions

The aim of this section is twofold. First, we review the literature concerning the relationship between the concept of genre, and its standardization or variability in legal discourse. We also review literature on contract language, though it is sparse from a genre theory perspective. Second, we review the particular legal and historical contexts of CWAs which make these documents specifically interesting for a study on the variability of contracts.

2.1 Legal genres: standardization or variation?

In this article, we approach genre with the conceptual framework created by researchers in language for specific purposes (Swales 1990, Bhatia 1993, Hopkins/Dudley-Evans 1988). Bhatia (1993: 13) defines genre as "a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs. Most often it is highly-structured and conventionalized with constraints on allowable contributions in terms of their intent, positioning, form and functional value." In this tradition, genres are stabilized communicative endeavors within a community of professionals. Genres, therefore, presuppose a degree of standardization that makes them culturally recognizable and judged as normatively legitimate by the members of the professional community itself.

Legal genres in particular face pressure to become and to remain standardized. They must be considered well-formed according to the legal standards of the jurisdictions and the legal authorities under which they are considered instruments. Legal texts that vary from an established legal genre may be judged void of legal force or can lead to sanctions by judicial authorities. In this way, legal genres tend to be distinct not only from non-specialized language, but also from other specialized genres (Gotti 2012: 52). This additional pressure towards standardization and risk-aversion leads to the oft-cited conservativeness and fixedness of legal documents (Gotti 2012: 52, Hiltunen 2012: 39). Indeed, innovation increases the risk of ambiguity and misinterpretation by stakeholders of legal genres. To avoid these risks, researchers argue, legal professionals aim for adhesion to approved legal genres (Gotti 2012: 52).

However, variation should be recognized as an equally necessary component of legal genres. Genres such as contracts are also instruments for accomplishing goals in dynamic

and long-term business relationships. As such, they must vary to meet the needs of the actors who employ them. This has not been stated clearly enough in the literature about legal genres, including contracts. While the literature highlights variability when comparing different legal genres or across legal systems (Trosborg 1997, Tiersma 1999, Bhatia 2012, Goźdź-Roszkowski 2012, Sun/Cheng 2017), there is little literature about variability within one genre, one language and one legal system, with the exception of Groom/Grieve's study of British patents (2019).

Linguistic studies on contractual language and its variability are, therefore, lacking. Most current literature about contractual language can be divided into two types: 1) prescriptive material, such as manuals and guides for professionals who draft and interpret contractual language (Chesler 2009, Bugg 2016, Cartwright 2016); 2) stylistics literature that is often focused on the question of simplifying contractual language (Richard 2021). Neither of these fields takes a genre-based approach as we do here. In the sparse linguistic studies, which deal directly with contracts as a genre, contracts have been represented as a "frozen" legal genre (Danet 1980: 471). Trosborg (1997) compares the pragmatics of contract language to other legal genres, also cites Danet (1980). From a pragmatic approach to genre, Trosborg (1997) argues that both the parties (exoteric audience) and their lawyers (esoteric audience) are receivers of contractual language. Trosborg further emphasizes the importance of the lawyers and the legal framework in which the contract functions:

> The function of the intermediary filter of the communicative process – the lawyer – is to ensure that the contents of the contract are in agreement with the scope for dispositional freedom on the part of the parties to the contract as limited by the legal framework. In this respect, legal discourse within the scope of language of the law can be described as formula-based communication in as regards form: the legal rules and their accumulated precedents function as a sort of matrix or mold, so that the form side of contracts is, in effect, determined normatively. For this reason, it is claimed that the form side is mainly oriented towards the esoteric audience – jurists and lawyers. (Trosborg 1997: 61)

Trosborg's (1997: 61) claim raises the question of how variation in the "matrix or mold" that is the legal framework would affect the "form" of a contract in terms of genres.

CWAs provide an opportunity to address this gap. Container liner shipping is characterized by a fluctuating legal framework and the need for strategic adaptation in a high-risk sector (OECD 2015). These conditions affect CWAs, which are the contracts that govern this industry. Returning to Trosborg's claim, we ask if and how a variable legal framework will affect this contractual genre, using lexical content as a proxy for genre. In addition, despite their enormous impact on the world economy (UNCTAD 2021), these contracts have not yet been the subject of linguistic analysis. There is, consequently, little information about their content, nor about how variation of these features interacts with the fragmented legal framework in which they function. We describe this framework in the following section.

2.2 Legal background of CWAs

This section describes the specific legal framework and historical context of liner shipping and of the related contracts in order to understand the impact this context has on the evolution and current function of CWAs. According to the World Shipping Council (WSC),

> Liner shipping is the service of transporting goods by means of high-capacity, ocean-going ships that transit regular routes on fixed schedules. There are approximately 400 liner services in operation today, most providing weekly departures from all the ports that each service calls […] (WSC 2022).

With regard to competition rules, liner shipping benefits from a derogatory regime and the sector is by nature highly-standardized around the globe (OECD 2015). While some model contracts such as CWAs exist, the rules applied are complex and unharmonized, which make this sector a unique context for studying the variation of the CWAs governing it.

2.2.1 Derogatory legal framework

Since its development in the last quarter of the 19th century, the liner shipping industry has been characterized by different forms of close cooperation between ocean carriers both at commercial and technical levels (Brooks 2000, Stopford 2009, Fedi 2021). This collaboration has been materialized and enforced through CWAs that evolved over time. Crucially, these agreements have been exempted from competition law (OECD 2015) and thus, their legal framework has always been derogatory regarding antitrust principles. In most countries, a block exemption regime is still granted to liner companies while monitored by competition authorities pursuant to different regulations (Fedi 2019, ITF 2019).

Historically, conferences and rate discussion agreements allowing price-fixing have led the contractual relationships of shipowners since the early beginning of the liner shipping development (circa 1870) till the nineties in the 20th century. Their aim was to limit a price war between maritime companies and to adjust transport capacities. Even though these types of agreements are still in force (ITF 2018, Fedi 2019), with the boom of containerization, they have progressively lost their dominance in favor of more technical CWAs. The latter agreements, such as 'consortia' and 'alliances,' mutualize and rationalize means of production while also minimizing the risks between shipowners (Fedi/Tourneur 2015, OECD 2015). Furthermore, the European Union (EU) prohibited maritime conferences as of 18 October 2008 (EU Council 2006, Fedi/Besancon 2008). The EU decision, however, had little impact (FMC 2012) and the conferential system has remained in many countries (Fedi 2013, Merk/Teodoro 2022).

To summarize, notwithstanding a strict compliance regime in some countries like the United States, liner shipping continues to evolve under a specific legal framework in the 21st century. However, this regime raises many questions particularly concerning competition law (Corruble 2018, Fedi 2019, 2021). The container liner industry is especially problematic. It is worth noting that the current contracts organizing the liner services bind direct competitors that are increasingly horizontally and vertically concentrated (Cariou 2008, ITF 2018). In 2020, more than 80 % of containers were carried by the ten leading maritime companies and the 'big four', i. e., Maersk, MSC, CMA CGM and COSCO represented around 58 % of this total volume (Alphaliner 2020). The three main alliances (2M, Ocean Alliance, THE Alliance) owned more than 90 % of market shares in 2018 (ITF 2018). Some voices, therefore, are calling for the modification of the current legal framework since it enables the most powerful companies to become larger (ITF 2018) and strengthen their bargaining power to the detriment of customers and suppliers as the COVID-19 crisis clearly showed (UNCTAD 2021).

2.2.2 Lack in law regulating container liner industry

Paradoxically, while container liner shipping is one of most globalized and standardized industries, its legal context is variable. There is neither an international convention setting out uniform provisions for the key clauses generally stated in the most common contracts nor a supranational institution in charge of the monitoring and or the settlement of disputes related to those contracts (Fedi/Tourneur 2015). The current legal framework applicable to CWAs is fragmented within different national or regional legislations (Fedi 2019). Consequently, as illustrated by the famous P3 Network, the same contract can receive a different interpretation depending on the competent competition authority (Corruble 2015, Fedi/Tourneur 2015). The P3 was approved by the EC and the FMC. It was, in contrast, refused by the Chinese competition authority (Braakman 2013). Even though this case highlighted the lack of uniform rules applied to CWAs (Fedi/Tourneur 2015), the shipping companies in question easily overcame the Chinese refusal by signing other contracts with additional partners. Finally, as merchants, liner companies have broad contractual freedom in the negotiation of their CWAs despite using model contracts. In this respect, we can underline that the principle of party autonomy attached to the theories of contract law is recognized in the context of international trade (UNIDROIT 2010: chap.1, art.1.1, Bix 2012). Fedi et al. (2022) have recently underlined that the existing CWAs entitle sea carriers to freely manage their activities with wide powers and ultimately to satisfy their needs. Obviously, the related parties benefit from broad contractual freedom providing them significant rights to achieve their objectives in terms of cost optimization, economies of scale and risk sharing. Moreover, the current digitalization trend prevalent in this sector has encouraged the adoption of digital cooperation contracts (e. g. DCSAA and Tradelens) facilitating the interoperability of information technology (IT) systems and raising questions on the exchange of sensitive information between members (Braakman 2020).

Notwithstanding abundant literature on transport economics and management of liner shipping (notably Brooks 2000, McLellan 2006, Cariou 2008, Luo/Fan/Wilson 2014), few scholars have paid attention to the content of the contracts governing the carriers' relationships so far (Corruble 2018, Fedi 2019, 2021). This research is crucial as strong criticism of shipping alliances and their possible strategic use of CWAs has recently emerged. A 2018 report from the International Transport Forum (ITF) underlined the negative effects of alliances beyond the significant market concentration, especially with regard to lower service frequencies, declining schedule reliability, longer waiting times, concentration of port networks, overcapacity, and barriers to entry into some trades (ITF 2018).

2.2.3 CWAs in container liner shipping

In terms of legal genres, the CWAs signed by container liner companies are private contracts that are not released to customers or suppliers of the co-contracting parties nor to any third parties. Crucially, CWAs are considered technical and not commercial contracts by international authorities and institutions (ITF 2018, 2019, EC 2020, Fedi et al. 2022). With the notable exception of the United States, the limited access to these contracts is linked to the sensitive information they contain and that liner companies are reluctant to divulge. It certainly justifies why the existing legal and linguistics literature merely refers to the main CWAs such as VSAs or SCAs but provides no description or analysis of their content. In this way, these instruments have been the subject of few studies. In addition, these contracts are diverse, especially in con-

tainer liner shipping (Fedi 2021). Whereas contracts covering liners that operate in American jurisdictions are published by the FMC (2022a), public access to these contracts has not been widely known and never exploited for a linguistic study (Lavissière/Fedi 2021).

In order to explain the relationship between variability, contractual genres, and legal context in the case of CWAs, we formulate three research questions (RQ) as follows:

RQ1: Are CWAs variable as a genre?

RQ2: Are there discernible trends in this variation?

RQ3: Can the variation be explained in terms of CWAs' fragmented legal framework?

3 Corpus and methodology

We chose to study variation in the CWA genre through the proxy of lexical words and textometry. The link between patterns in lexical content and genre was shown in research by Biber/Connor/Upton (2007). These authors take this approach to studying rhetorical structure divisions (*moves*) in genre. They use Vocabulary-Based Discourse Units (Biber/Connor/Upton 2007: 17). These are a group of words that appear frequently together in a given type of document. Vocabulary-Based Discourse Units allow researchers to divide a document into coherent units according to lexical trends.

The textometric software that we use to analyze the CWAs also works on the principle of finding and representing lexical words that appear in the same type of unit of discourse. In our case, the unit of discourse is smaller and generally corresponds to a sentence. This unit is relevant for law as legal clauses are often contained in one sentence, although sentences in legal documents tend to be long (Hiltunen et al. 2001: 56). Reinert (1983: 189–192), whose algorithm we use in Iramuteq, argued that utterances or sentences ("énoncés") bear the subjective mark of their utterers who make a statement about the outside world by putting lexical words together in the form of an object (the subject of the sentence) and a statement about the object (the predicate). When two lexical words appear in the same unit of context, Reinert (1990: 31) calls this a "local representation". When two words appear in a unit of context multiple times in a corpus, Reinert (1990: 31) calls this a "lexical world": "a set of statements that show the same global perception of a world."[1] He operationalizes the concept of 'lexical world' through a descending hierarchical classification (DHC) algorithm, which we describe in section 3.2.2.

In sum, this method was chosen because it can show overall trends in variation of lexical content of different subtypes of documents and in time. In Reinert's (1990) theory, this variation corresponds to different representations of the world within a corpus. It descends from the school of Geometric Data Analysis (Le Roux/Rouanet 2005), initiated by Jean-Paul Benzécri (Beaudouin 2016). While some corpus linguists have criticized this method as descriptive, it aims at making underlying lexical trends visible for qualitative interpretation (Beaudouin 2016). This section is divided into two parts. We first describe our corpus. Then we describe our method as applied through Iramuteq.

[1] This is our translation of "un ensemble d'énoncés connotant une même perception globale d'un monde".

3.1 Corpus

Our corpus consists of 46 agreements collected from the official website of the FMC (FMC 2022a).[2] We included all the CWAs that were available between May and September 2020 and that were classified by the FMC as Vessel Sharing Agreement (VSAs), Slot Charter Agreement (SCAs), Alliances, and CWAs without a secondary classification. They represent CWAs published between 1985 and 2020. All the agreements included in our corpus have the term *Cooperative Working Agreement* on their official title page. The total number of words in the corpus is around 88,140.[3]

While all of the agreements included are CWAs according to their official title page, some also have a secondary classification as VSAs, SCAs, or Alliances by the FMC. We only included the contents of the articles of these agreements. Their title pages, table of contents, signature page(s), and appendices were excluded. No republications of agreements were incorporated in order to avoid introducing the bias of similar republished texts into the corpus because republications may contain the same wording, while simply introducing a new partner into the agreement.

3.2 Methodology

Our methodology included two steps. The first step was the preparation of the corpus, and the second was a textometric analysis using the software Iramuteq (Ratinaud 2014). The corpus was prepared in two stages. First, the documents were manually converted into a text file, and the text file was checked for accuracy. Second, we coded each agreement for type on the official title page (TYPE), for second classification (TYPE2), for title (TITLE), for agreement number (NUMBER) and for effective date, in terms of the year the agreement became effective (DATE).

To analyze the content of the articles themselves, we carried out a textometric analysis using the software Iramuteq (Ratinaud 2014). The software goes through preliminary steps before performing selected algorithms. First, upon importation, the software divides the corpus into units of context[4]. The length of the units of context is determined by the researcher during the importation stage (Loubère/Ratinaud 2014). In this study, the length of the unit of context was set to 45 words, which corresponds to the average length of sentences in certain legal texts (Hiltunen et al. 2001: 56). Second, Iramuteq lemmatizes the text and divides it into three categories of *forms*: *active* (henceforth, *lexical words*), *supplementary* (henceforth, *grammatical words*), and *hapax legomenon* (henceforth, *hapax*). Lexical words include adjectives, adverbs, verbs, and nouns. Grammatical words include prepositions, pronouns, determiners, conjunctions, and other grammatical elements. Hapax includes all words that appear only once in the corpus. The categorization is carried out according to the software's dictionaries (Loubère/Ratinaud 2014).

---

2 A complete list of the agreements and their variables, including secondary classification and effective date, can be provided upon request to the authors.

3 Some of the agreements are short, around 2000 words. These are full contracts. They usually concern slot sharing on vessels.

4 These are called *text segments* in Iramuteq's terminology. We use the term *unit of context* in this article because we find that it more clearly expresses the concept behind the term as theorized by Reinert (1983).

After these steps, we used the DHC proposed by Reinert (1983) and modified for Iramuteq (Marchand/Ratinaud 2012). The software creates a contingency table with units of context in rows and lexical words in the columns. It uses presence (1) or absence (0) as criteria for filling the table. Iramuteq then carries out a correspondence analysis using the contingency table. The first division in the table is made according to maximum inter-class inertia. The DHC algorithm then continues dividing the table iteratively until further division does not increase inter-class inertia. Then, chi-squared values are used to determine which lexical words are specific to each class and to assign them accordingly. The classes produced by the DHC are interpreted in the following section. Finally, the software also allows for comparing the association strength of each variable in comparison with the classes resulting from the DHC. We compared the association of the variables DATE and TYPE2 with the classes produced by the DHC.

4 Results

The following section describes the four classes constructed by the DHC and the trends in variation that we observed thanks to the textometric analysis.

4.1 Classes

The DHC results in four classes which we named *Business framework*, *Strategic exchanges*, *Technical exchanges* and *Legal framework*. It should be highlighted that these classes, or lexical worlds, were constituted by the software through statistical methods and were not created by the authors. The authors interpret the results of the statistical analysis by assigning coherent names to the classes, as can be seen in Figure 1 on page 139.

In our corpus, the first division is made between the classes containing lexical words that refer to what is exchanged through the agreement and the frameworks within which this exchange takes place. We provide an overview of the lexical words in these four classes and the characteristic units of context in the following paragraphs. We also include most characteristic units of context, as determined by the software. The examples of lexical words and units of context provided below are all significantly associated with their classes according to Pearson's chi-squared test ($p = <.001$).

First, the class *Business framework* includes 43 % of the corpus, the largest percentage of the corpus. This is a surprising finding because CWAs are technical contracts and several jurisdictions forbid commercial exchanges between the parties. However, CWAs concluded over the last decade encompass the entire maritime supply chain and involve a door-to-door collaboration in all transport modes and interfaces (warehouses, logistic platforms or seaports). This may partially account for the continuing dominance of this class. This class includes lexical words related to formal communication, such as *notice* and *write*. It also includes lexical words referring to the limits of the agreement in time*, effective, date, day, terminate, termination*. While many contracts contain these words, recent CWAs have been concluded without an expiration date or renewed indefinitely by tacit acceptance as seen in example (1) below (emphasis added by the authors):

(1) This agreement shall remain in effect until *terminated* by unanimous consent of the Parties or until all but one Party has withdrawn.

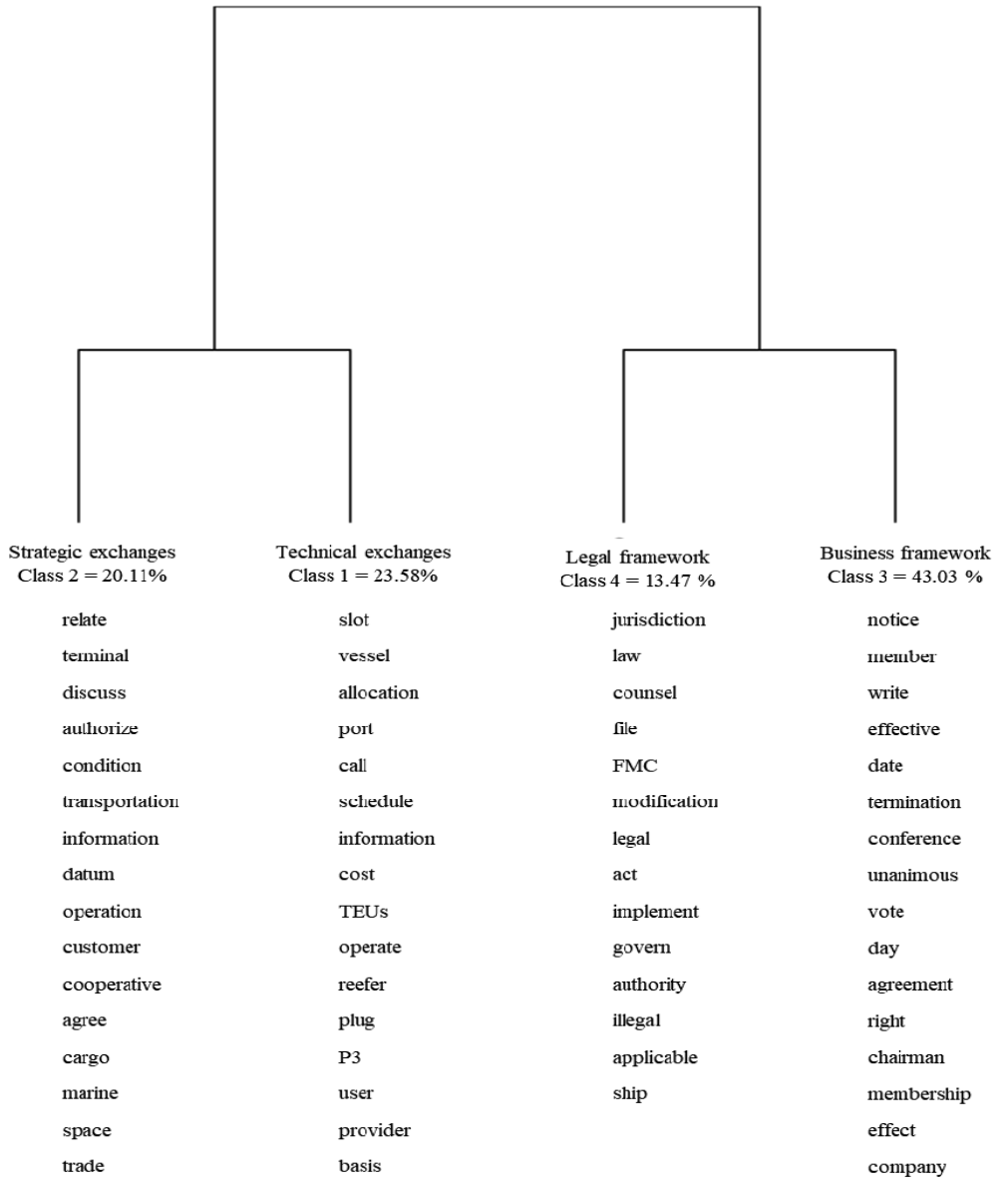

Figure 1: Four classes resulting from the DHC analysis

Furthermore, *Business framework* includes lexical words referring to the parties to the agreement and their relationship within the agreement. These include *member*, *conference*, *vote*, *unanimous*, *chairman*, *right*, *membership*, *company*, as seen in (2):

(2) The active *members* are authorized to form, own, operate, and dissolve the Digital Container Shipping Association […] as a separate legal entity under Dutch law to carry out the authorities set forth herein and are authorized to discuss and agree on all aspects of the structure.

Second, the class *Strategic exchanges* includes 23 % of the corpus and is the second largest class. Many of the units of context in this class come from the article *Agreement Authority,* which is the most complex article and sets out a wide field of action for the parties. It includes lexical words referring to collaborative action, such as *relate*, *discuss, authorize*, *agree*, *cooperative*, as seen in (3) below:

(3) The parties are *authorized* to exchange information on any scope of this agreement and to reach agreement on any and all administrative functions related hereto including but not limited to forecasting, terminal planning, insurance liability, cargo claims, indemnities.

This class also includes units of context referring to some limits on exchanges, showing that the contracts discursively distance themselves from being commercial. This is especially true of more recent contracts as in (4), which was approved by the FMC in 2015:

(4) Each party shall conduct its own separate *marketing* and sales activities, shall issue its own bills of lading, and unless otherwise agreed, handle its own claims […].

This class includes lexical words related to strategic gains from the relationship established by the CWAs. It confirms the latent objectives of the contracting parties as regards rationalization, optimization, risk sharing and economic efficiency to be sought, including words such as *terminal*, *operation*, *marine*, *trade*, *transportation*, as seen in (5):

(5) The parties are authorized to discuss and agree upon arrangements for the use of *terminals* in connection with the chartering of space hereunder including entering into exclusive preferential or cooperative working arrangements with marine *terminal* operators and other persons relating to marine *terminal* stevedoring or other shoreside services.

Finally, the class includes lexical words related to the actual exchanges*, cargo*, *space*, *datum*, *information,* and *customer*. This includes units of context referring to the exchange of these strategic resources as well as to certain limits, as seen in (6) below:

(6) Notwithstanding the preceding sentence, no information which is commercially sensitive (*customers* (save as necessary to comply with the terms of a particular contract of carriage), *customer* pricing and other, similar commercially sensitive information) may be exchanged hereunder directly or indirectly between any of the Parties.

It should be noted, however, that this example comes from the P3 VSA agreement, an alliance which was judged as in violation of Chinese competition law, despite being first approved by the FMC.

Third, the class *Technical exchanges* represents around 20 % of the corpus. This class incorporates the lexical words referring to the technical aspects of the terms and gains for each party. This covers lexical words related to vessel chartering, such as *slot, space, allocation, cost, teus (twenty equivalent units i.e. a container)*. It also includes lexical words referring to the sharing of vessels and other material, *vessel, capacity*, *reefer plug,* and to the actors and their actions, *user*, *provider, operate, agree*. Finally, this class incorporates lexical words related to the routes of the ships, *call*, *port*, *schedule*, such as in (7):

(7) The parties are authorized to discuss and agree on the *ports* to be called, *port* rotation, itineraries, service speed, and all other aspects of the structure and scheduling of the services to be operated hereunder.

Interestingly, this class is statistically associated with the article entitled *Agreement Authority*, rather than with the more technical articles such as *Geographic Scope* or *Purpose*. We interpret this to mean that technical exchanges are highly related to the strategic exchanges and the market, despite clauses such as (4) above which declare otherwise.

Fourth, the class *Legal framework* represents around 13 % of the corpus, the smallest percentage. This class integrates words related to the actors, such as the *Federal Maritime Commission, authority*, *implement*, *ship*, and *counsel*. It also includes the actions that are carried out concerning the agreement and its parties, *file* and *govern*. Finally, it refers to legal frameworks and instruments: *jurisdiction, law, legal, act,* and *applicable*.

4.2 Variation according to type and to date

In terms of the variation in CWAs, the DHC shows that there is significant variation between the contracts according to the variables TYPE2 and DATE. We observe that certain classes are more highly associated with the different categorizations on the website of the FMC (2022a), or with different effective dates.

Figure 2a below shows a graph of the Chi-squared values for significant associations ($p < .001$) between TYPE2 and the DHC classes. The class *Technical exchanges* is significantly associated with SCAs, VSAs and Alliances. It is negatively associated with CWAs with no secondary classification. *Business framework* is significantly associated with CWAs that do not have a secondary classification. *Legal framework* is also significantly associated with CWAs whose secondary classification by the FMC is VSA.

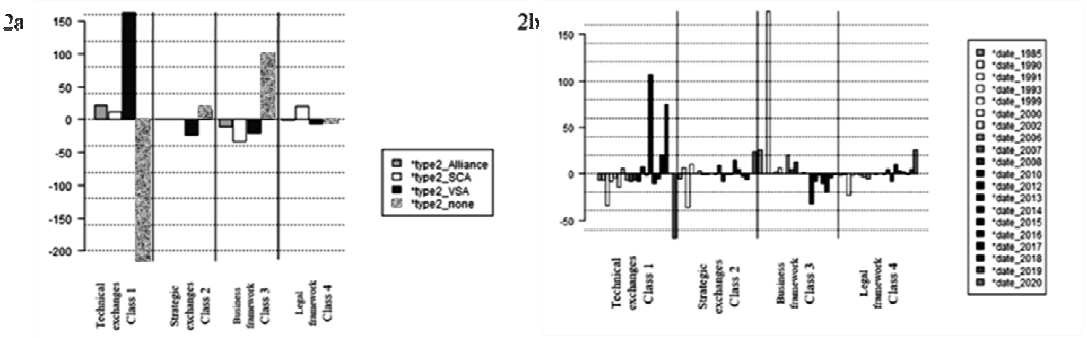

Figure 2a and 2b: Chi-squared values for the associations between TYPE2 (2a) and DATE (2b) and the DHC classes

However, as we highlighted in section 4.1, the units of context that contain language expressing *Technical Exchanges* are statistically associated with the same article as those in the *Strategic Exchanges* class. We interpret this as more technically-oriented contracts aiming for the same strategic goals as those that are business-oriented. This was the conclusion of the Chinese competition law authority concerning the P3 Alliance in 2014.

In terms of diachronic variation, the Chi-squared values indicate significant associations

between the DHC classes and the variable DATE. This is shown in figure 2b. The class *Technical exchanges* has higher Chi-squared values in association with more recent CWAs. This constitutes a trend in the industry: opting for more technical agreements, probably caused by the greater complexity of end-to-end services provided by the parties and also by the pressure from several different competition law authorities in reaction to the effects of the contracts on the market. *Legal framework* also shows slightly higher Chi-squared values in more recent CWAs, perhaps another reaction to increasing pressure to comply with competition law. The class *Business framework* shows higher Chi-squared values with older CWAs. The Chi-squared values of *Strategic exchanges* varies in time, but shows a slight overall positive trend. It is important to note that the change towards more technical language began around 2014, the year when the P3 Alliance failed to gain approval from the Chinese competition law authority.

In sum, CWAs vary in terms of the amount of language dealing with either the technical nature of the cooperation or the business structure created through the contract. We observe an overall trend towards more technical lexical items, probably in reaction to increasing complexity of door-to-door transport service provision and stricter monitoring from competition law entities. Given that there is no one single authority, however, important exceptions to this trend exist, such as the DCSAA.

5 Discussion

Our results make several contributions to the literature about contracts, which from a linguistics and genre analysis perspective, is scarce. In terms of contractual language, the literature is limited to prescriptive guidelines about how to draft or interpret contracts for legal professionals (Chesler 2009, Bugg 2016, Cartwright 2016) or about simplifying contractual language in the wake of the Plain Language movement (Richard 2021). We contribute to this literature by describing the actual lexical content of a specific type of technical contracts, CWAs. These contracts have not been subject to a linguistic or legal description despite their importance for the industry and for individual consumers. While this is only a first step towards filling the larger gap of defining contracts as a genre, we show that CWAs have four lexical worlds: *Business framework*, *Strategic exchanges*, *Technical exchanges* and *Legal framework*. These results led us to answer our RQs in the paragraphs below.

First, in RQ1, we asked whether CWAs are a "frozen" genre in the larger panorama of contracts (Danet 1980). Our results show that, on the contrary, CWAs vary substantially. While all four of these lexical worlds exist in each contract, they are not equally represented if we compare the lexical content to the FMC's categorization (TYPE2). VSAs, SCAs, and Alliances tend to be more clearly associated with technical exchanges. However, we cannot simply assume that this variation is due to evolving technical needs in the industry rather than commercial stakes. As our results demonstrate, the technical language replaces business framework language in the article *Agreement Authority* rather than increasing the length of articles dedicated to technical subjects.

Second, we asked whether there were discernible trends in CWAs' variation (RQ2). We observed a clear trend towards more technical lexical items in time. The lexical items more closely associated with older documents focus on the members and their relationship in activities such as voting and being a shareholder. In more recent contracts, the focus is on the specifics of the technical exchanges, such as the shipping capacity, which one company will put at the disposal of another. This more technical bend in the words used in the agreements be-

gins around 2013–2014, the period in which the 3P Alliance was declared in breach of Chinese competition law (Braakman 2013). However, this trend has many exceptions. CWAs with no secondary classification have also been published recently. In line with Fedi (2021), we observe that recently filed CWAs, such as the DCSAA and Tradelens, may diverge widely from the general trend towards contracts with more technical lexical items that deal with the digitalization strategies set up by the maritime companies. The management of IT systems, electronic bills of lading, blockchain, internet of things, data security or software holds a central place in this new type of CWA, which covers end-to-end services (Braakman 2020).

Third, we show that the fragmented legal framework in which CWAs exist allows them to take on multiple forms, such as SCAs, VSAs, Alliances, or newer forms that have not yet been named by the FMC (RQ3). The DHC reinforces our interpretation that the fragmented framework allows agreements to change their linguistic approach to reaching the same strategic goals for the parties while trying to avoid sanctions from authorities. The agreements not only set up technical exchanges of slots or of vessels, but they also lay the framework for business cooperation. The dominance of the *Business framework* and the presence of *Strategic exchanges* confirms the relevance of the legal literature classifying these documents as problematic: CWAs create merger or joint venture-like frameworks. As a result, they challenge competition law and call into question the relevance of the current exemption rules (Corruble 2018, Fedi et al. 2022). The latest digital CWAs that aim to share interoperable IT standards along the international transport supply chain (e.g. DCSAA) also reinforce this question on the suitability of US and EU rules (Braakman 2020). Moreover, CWAs ensure more rights than obligations and appear as the expression of an unilateral power, a contractual trend highlighted notably by Lokiec (2021) and Ancel (2008).

In sum, the fragmented legal framework allows CWAs to be dynamic and polymorphic legal instruments. They adapt to new legal restraints by changing their lexical tone and to operational constraints by enlarging the purpose of CWAs. For this reason, they are increasingly used by maritime shipping lines to reach goals such as terminal integration, equipment optimization, economies of scale, data sharing, IT development or customer exchange. These companies can exploit the variability of the genre to create contracts such as the recent digitally-oriented CWAs that meet goals beyond merely sharing technical capacities. We conclude this variation of CWAs is justified by the current fragmented legal framework applicable to the sector and the leading companies, which take great advantage of the lack of legal "mold" to constrain the CWA genre. This also confirms the "degree of opportunism" shown by the parties of cooperation agreements as demonstrated by Axerold (1984) in the 1980s.

6 Conclusion

From legal and linguistic perspectives, the fragmented legal framework has allowed CWAs to evolve over time and to avoid new legal constraints in the liner shipping industry. Our results show that the CWAs vary lexically in time and according to the secondary classification by the FMC. This variation gives evidence for Bhatia's (2012) argument that genres do not need to exhibit full standardization in order to act as instruments for given professional communities to reach their goals. The fragmented legal framework in which CWAs currently exist allows much margin for evolution rather than the firm constraining mold described by Trosborg (1997).

With regard to our future research agenda, more research should be done on CWAs both

from a linguistic and legal perspective. From a linguistic point of view, the articles in CWAs should be studied and the date of their appearance analyzed. The articles should also be further divided into legal clauses and their position analyzed in time. From a legal perspective, considering the lack of coherence between some of the CWAs' titles and their content, such as the DCSAA, we will propose a more detailed study of the legal nature of clauses contained in CWAs. In line with preliminary research (Fedi 2021), this classification would facilitate the knowledge and contribute to a better understanding of CWAs in container liner shipping. Further, a comparison between the language of CWAs and that of other legal documents used in business collaboration such as joint ventures will be necessary to understand the role of CWAs in the context of increasing concentration in the liner shipping industry (ITF 2018).

References

Alphaliner (2020): *Alphaliner Top 100 on 31 December 2020.* https://alphaliner.axsmarine.com/PublicTop100 (27.06.2022).

Ancel, Pascal (2008): "L'accroissement de l'unilatéralité dans le contrat?" *La contractualisation de la production normative:* 29-44. Eds. Sandrine Chassagnard-Pinet / David Hiez. Paris: Dalloz.

Axerold, Robert (1984): *The Evolution of Cooperation*. New York: Basic Books.

Beaudouin, Valérie (2016): "Statistical Analysis of Textual Data: Benzécri and the French School of Data Analysis." *Glottometrics* 33, 56–72. https://hal.telecom-paris.fr/hal-02191472 (15.05.2022).

Biber, Douglas / Connor, Ulla / Upton, Thomas Albin (2007): *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure.* Amsterdam/Philadelphia: Benjamins.

Bhatia, Vijay K. (1993): *Analysing Genre: Language Use in Professional Settings*. London: Longman.

Bhatia, Vijay K. (2012): "Critical Reflections on Genre Analysis." *Iberica* 24: 17–28.

Bix, Brian H. (2012): *Theories of Contract Law*. Cambridge: Cambridge University Press.

Braakman, August J. (2013): "The Anti-competitive Effects of a Globally Concentrated, Oligopolistic Maritime Market: from Explicit to Tacit Collusion – an Analysis Based on P3 Network." *Journal of International Maritime Law* 19.5: 419–436.

Braakman, August J. (2020): "Does the Digital Leviathan Spell the End of Fair Competition in the Shipping Industry?" *Journal of International Maritime Law* 26: 1–8.

Brooks, Mary R. (2000): *Sea Change in Liner Shipping: Regulation and Managerial Decision-Making in a Global Industry*. Oxford: Pergamon.

Brooks, Mary R. / Meersman, Hilde / Sys, Christa / Van de Voorde, Eddy / Vanelslander, Thierry (2019): *Technical Report: Regulation in the Liner Shipping Industry: Pathway to a Balance of Interests*. TPR Department of Transport and Regional Economics, University of Antwerp.

Bugg, Stuart G. (2016): *Contracts in English: An Introductory Guide to Understanding, Using and Developing "Anglo-American" Style Contracts*. München: Beck.

Cariou, Pierre (2000): *Les alliances stratégiques dans le transport maritime de lignes régulières: efficacité ou pouvoir de marché*? Thèse Nantes, University of Nantes.

Cariou, Pierre (2008): "Liner Shipping Strategies: An Overview." *International Journal of Ocean Systems Management* 1: 1–7.

Cartwright, John (2016): *Contract Law: An Introduction to the English Law of Contract for the Civil Lawyer.* Oxford, UK / Portland: Hart Publishing.

Caschili, Simone / Medda, Francesca / Parola, Francesco / Ferrari, Claudio (2014): "An Analysis of Shipping Agreements: The Cooperative Container Network." *Networks and Spatial Economics* 14.3: 357–377. https://doi.org/10.1007/s11067-014-9230-1 (24.08.2022).

Chesler, Susan M. (2009): "Drafting Effective Contracts: How to Revise, Edit, and Use Form Agreements." *Business Law Today* 19.35: 35–37.

Corruble, Philippe (2015): "Global Shipping Alliances: What Prospects after the Rejection of the P3 Project by China's Competition Authority?" *Journal of International Maritime Law* 21.3: 203–211.

Corruble, Philippe (2018): "Les alliances maritimes mondiales entre le zist et le zest." *Droit Maritime Français* 807: 867–876.

Danet, Brenda (1980): "Language in the Legal Process." *Law & Society Review* 14.3: 445–564.

European Commission (EC) (2020): *Commission Regulation (EU) 2020/436 of 24 March 2020 amending Regulation (EC) No 906/2009 as regards its period of application* (Text with EEA relevance) C/2020/1734. OJ L 90. (25.03.2020): 1–2.

European Council (EU Council) (2006): *Council Regulation (EC) No 1419/2006 of 25 September 2006 repealing Regulation (EEC) No 4056/86 laying down detailed rules for the application of Articles 85 and 86 of the Treaty to maritime transport, and amending Regulation (EC) No 1/2003 as regards the extension of its scope to include cabotage and international tramp services* (Text with EEA relevance). OJ L 269. 28.9.2006: 1–3.

Fedi, Laurent (2013): "L'abrogation des conférences maritimes dans l'Union européenne: premiers bilans et perspectives d'évolution." *Droit Maritime Français* 750: 697–711.

Fedi, Laurent (2019): "Alliances stratégiques dans l'industrie des liners: les voies de réponses du droit de la concurrence européen et international." *Droit Maritime Français* 816: 675–686.

Fedi, Laurent (2021): "Les contrats d'alliances maritimes: Coopération intégrée ou fusion masquée?" *Penser le Maroc logistique: une perspective pluridisciplinaire.* Eds. Abdelali Naciri / Gilles Paché / Michel Savy. Marseille: Presses universitaires d'Aix-Marseille. 161–173.

Fedi, Laurent / Besancon, Marion (2008): "La fin du régime des conférences maritimes: vers une concurrence pure et parfaite du transport par mer au départ ou à destination de l'Union européenne?" *Droit Maritime Français* 696: 791–803.

Fedi, Laurent / Tourneur, Margaux (2015): "Les consortia et les alliances (géo)stratégiques face aux nouveaux enjeux du transport maritime conteneurisé." *Droit Maritime Français* 769: 385–399.

Fedi, Laurent / Faury, Olivier / Rigot-Muller, Patrick / Nicolas, Montier (2022): "COVID-19 as a Catalyst of a New Container Port Hierarchy in Mediterranean Sea and Northern Range." *Maritime Economics and Logistics: Special issue: Maritime and Port Governance.* Ed. Hercules Haralambides. https://doi.org/10.1057/s41278-022-00223-z (24.08.2022).

Federal Maritime Commission (FMC) (2012): *Study of the 2008 Repeal of the Liner Conference Exemption from European Competition Law*. Washington, DC: Federal Maritime Commission.

Federal Maritime Commission (FMC) (2022a): *Federal Maritime Commission Agreement Library*. https://www2.fmc.gov/FMC.Agreements.Web/Public (23.08.2022).

Federal Maritime Commission (FMC) (2022b): *FMC Approves 2 Million Settlement Agreement with Hapag Llyod*. June 8, 2022. https://www.fmc.gov/fmc-approves-2-million-settlement-agreement-with-hapag-lloyd/ (15.05.2022).

Ghorbani, Mohammad / Acciaro, Michele / Transchel, Sandra / Cariou, Pierre (2021): "Strategic Alliances in Container Shipping: A Review of the Literature and Future Research Agenda." *Maritime Economics and Logistics* 24: 439–465. https://doi.org/10.1057/s41278-021-00205-7 (24.08.2022).

Gotti, Maurizio (2012): "Text and Genre." *The Oxford Handbook of Language and Law*. Eds. Lawrence M. Solan / Peter M. Tiersma. Oxford, UK: Oxford University Press. 52–66. https://doi.org/10.1093/oxfordhb/9780199572120. 013.0005 (24.11.2020).

Goźdź-Roszkowski, Stanislaw (2012): *Patterns of Linguistic Variation in American Legal English*. Bern: Lang.

Groom, Nicholas / Grieve, Jack (2019): "The Evolution of a Legal Genre." *Corpus-based Research on Variation in English Legal Discourse*. Eds. Teresa Fanego / Paula Rodriguez-Puente. Amsterdam: Benjamins. 201–234.

Hiltunen, Risto (2012): "The Grammar and Structure of Legal Texts." *The Oxford Handbook of Language and Law.* Eds. Lawrence M. Solan / Peter M. Tiersma. Oxford, UK: Oxford University Press. 39–51. https://doi.org/10.1093/oxfordhb/9780199572120.013.0004 (24.11.2020).

Hiltunen, Risto (2001): "'Some Syntactic Properties of English Law Language': Twenty-five Years after Gustafsson (1975)." *English in Zigs and Zags.* Eds. Risto Hiltunen / Keith Battarbee / Matti Peikola / Sanna-Kaisa Tansk. Turku: University of Turku. 53–66.

Hopkins, Andy / Dudley-Evans, Tony (1988): "A Genre-based Investigation of the Discussion Sections in Articles and Dissertations." *English for Specific Purposes* 7.2: 113–121. https://doi.org/10.1016/0889-4906(88)90029-4 (24.08.2022).

International Maritime Organization (IMO) (2020): *Introduction to IMO.* https://www.imo.org/en/About/Pages/Default.aspx (15.12.2020).

International Transport Forum (ITF) (2018): *The Impacts of Alliances in Container Shipping – Case Specific Policy Analysis*. Paris: ITF.

International Transport Forum (ITF) (2019): *Container Shipping: Data for the Evaluation of the EU Consortia Block Exemption Regulation, International Transport Forum*. Paris: OECD.

Justice Department (2022): *Justice Department and Federal Maritime Commission Reaffirm and Strengthen Partnership to Promote Fair Competition in the Shipping Industry.* February 25, 2022. https://www.justice.gov/opa/pr/justice-department-and-federal-maritime-commission-reaffirm-and-strengthen-partnership (24.08.2022).

Lavissière, Mary C. / Fedi, Laurent (2021): "Unfreezing Legal Genres: Maritime Shipping Agreements and Diachronic Variation". Conference Presentation, SLE Societas Linguistica Europeae Conference, SLE 2021, 30 August – 3 September 2021, Athens, Online Event.

Le Roux, Brigitte / Rouanet, Henry (2005): *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. Dordrecht: Springer. https://doi.org/10.1007/1-4020-2236-0.

Lokiec, Pascal (2021): *Contrat et pouvoir.* Paris: LGDG.

Loubère, Lucie / Ratinaud, Pierre (2014): *Documentation IRaMuTeQ 0.6 alpha 3*. http://www.iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf (24.11.2020).

Luo, Meifeng / Fan, Lixian / Wilson, Wesley W. (2014): "Firm Growth and Market Concentration in Liner Shipping." *Journal of Transport Economics and Policy* 48.1: 171–187.

Marchand, Pascal / Ratinaud, Pierre (2012): "L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011)." *JADT 2012. Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles.* Liège: Presses Universitaires de Liège. 687–699.

McLellan, Rogan (2006): "Liner Shipping Development Trends." *Maritime Policy & Management* 33.5: 519–525.

Merk, Olaf / Teodoro, Antonella (2022): "Alternative Approaches to Measuring Concentration in Liner Shipping." *Maritime Economics and Logistics: Special issue: Maritime and Port Governance.* Ed. Hercules Haralambides. https://doi.org/10.1057/s41278-022-00225-x (24.08.2022).

Ocean Shipping Reform Act (1998). 105th US Congress, Public Law, 105–258, October, 14. https://www.congress.gov/105/plaws/publ258/PLAW-105publ258.pdf (24.08.2022).

Organisation for Economic Co-operation and Development (OECD) (2015): *Competition Issues in the Liner Industry. Note by the Secretariat. Working Party No. 2 on Competition and Regulation*. DAF/COMP/WP2 2015.3.

Orts Llopis, María Ángeles (2014): "Contractual Commitment or Obligation? The Linguistic Interactions in Charter Parties Publication." *Interpersonality in Legal Genres*. Eds. Breeze, Ruth / Gotti, Maurizio / Sancho Guinda, Carmen. Bern: Lang. 87–113.

Ratinaud, Pierre (2014): *IRaMuTeQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*. Windows, GNU/Linux, Mac OS X. http://www.iramuteq.org (24.08.2022).

Reinert, Max (1983): "Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte." *Cahiers de l'Analyse des Données* 8.2: 187–198.

Reinert, Max (1990): "Alceste une méthodologie d'analyse des données textuelles et une application: Aurelia de Gerard de Nerval." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 26.1: 24–54. https://doi.org/10.1177/075910639002600103 (24.08.2022).

Richard, Isabelle (2021): "La 'dé-spécialisation' de l'anglais juridique: exemple de l'évolution de la stylistique contractuelle en Common Law." *Langues et langages juridiques.* Eds. Renaud Baumert / Albane Geslin / Stéphanie Roussel. Bayonne: Institut Francophone pour la Justice et la Démocratie. 143–160.

Shipping Act (1984): 98th US Congress, Public Law 98-237, March 20. https://www.govinfo.gov/content/pkg/STATUTE-98/pdf/STATUTE-98-Pg67.pdf#page=1 (24.08.2022).

Stopford, Martin (2009): *Maritime Economics*. 3rd edition. Oxon: Routledge.

Sun, Yuxiu / Cheng, Le (2017): "Linguistic Variation and Legal Representation in Legislative Discourse: A Corpus-based Multi-dimensional Study." *International Journal of Legal Discourse* 2.2: 315–339. https://doi.org/10.1515/ijld-2017-0017 (24.08.2022).

Swales, John (1990): *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

The White House (2021): "Executive Order on Promoting Competition in the American Economy." July, 9. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/07/09/executive-order-on-promoting-competition-in-the-american-economy/ (24.08.2022).

Tiersma, Peter M. (1999): *Legal Language*. Chicago: University of Chicago Press.

Trosborg, Anna (1997): *Rhetorical Strategies in Legal Language: Discourse Analysis of Statutes and Contracts*. Tübingen: Narr.

UNCTAD (2021): *Container Shipping in Times of Covid-19: Why Freight Rates Have Surged and Implications for Policymakers*. Policy brief n°84, April.

UNIDROIT (2010): *UNIDROIT Principles*. https://www.unidroit.org/instruments/commercial-contracts/unidroit-principles-2010/ (23.08.2022).

World Shipping Council (WSC) (2022): *About the Industry. How Liner Shipping Works*. *From There to Here*. https://www.worldshipping.org/how-container-shipping-works (23.08.2022).

Mary C. Lavissière, Ph.D., Associate Professor
Orcid: 0000-0 002-4241-4389
Nantes Université,
Centre de recherche sur les identités
les nations et l'interculturalité, CRINI
UR 1162, 44000 Nantes
marycatherine.lavissiere@univ-nantes.fr

Laurent Fedi, Ph.D.-HDR, Associate Professor
Orcid: 0000-0002-0332-2684
Kedge BS – Kedge Business School
CESIT – Maritime Supply Chain Lab
Domaine de Luminy
Rue Antoine Bourdelle, 13009 Marseille

Towards a Marine Accident Frame: Extraction of Predicate-Argument Structures with Word Sketches

Päivi Pasanen

Abstract This paper presents the results of a study that combines information extraction by means of word sketches with corpus-based analysis. In the analysis, terminological information was extracted from a corpus of investigation reports by retrieving the word sketches of the selected verbs. The aim of the study is to identify predicate-argument structures in a specific domain corpus by adopting the principles of Frame-based Terminology (FBT). Thus, this research shows how FBT can be applied to different scenarios such as a marine accident event, which is the object of this study. The long-term goal is to build a marine accident frame by utilising the predicate-argument structures identified in this study. The predicate-argument structures and the evolving frames can be used to enhance the representation and understanding of marine accident events. Thus, the predicate-argument structures are presented for consultation to future end users of the results, namely the stakeholders in the field of maritime safety, including maritime English teachers, as well as LSP teachers and terminologists.

Keywords corpus linguistics, Frame-based Terminology, Frame Semantics, marine accident event, predicate-argument structure

1 Introduction

Over the last twenty years, terminology has experienced a shift towards cognitive and corpus-based terminology. The shift is manifested in a cognitive approach that is known as Frame-based Terminology (FBT) (cf. Faber et al. 2006, Faber 2014: 14). FBT has adapted the basic principles of Frame Semantics (cf. Fillmore 1977, 1985, 2006 [1982], Fillmore/Atkins 1992) for structuring specialised domains and creating conceptual representations in specific domains (Faber et al. 2006: 192). In Frame Semantics, the key assumption is that the meanings of lexical units are constructed in relation to background knowledge, the structure of which is represented in semantic frames (cf. e. g. Fillmore 2006 [1982]: 378). The principles of FBT have been applied in a number of specialised domains for the creation of domain-specific frames and termbases; however, the maritime field has not been addressed in any depth. Hence, our study aims at introducing FBT in the maritime field by applying the principles and methods of FBT to one event concept, namely a *marine accident*. Here, a marine accident is an event that leads the vessel's operation from a safety sphere (state, situation) through a change towards an accident.

In texts, events are often represented linguistically in the form of predicate-argument structures (Faber 2014: 21). It is our aim to identify predicate-argument structures in a specific domain corpus and, furthermore, to describe the conceptual structures of the marine accident event and the lexical components in these structures. The evolving conceptual structures

Zitiervorschlag / Citation:
Pasanen, Päivi (2022): „Towards a Marine Accident Frame: Extraction of Predicate-Argument Structures with Word Sketches." *Fachsprache. Journal of Professional and Scientific Communication* 44.3–4: 148–168.

can be utilised for querying and organising data by researchers, terminologists and accident investigators, for example. We focus on the verbs used to express processes and activities, as well as to relate the participants involved and their roles. These participants are the semantic arguments of the predicates analysed in this study with each activating a certain argument structure. The long-term goal is to build a marine accident frame, and consequently, enhance the representation and understanding of marine accident events.

Shipping is a complex system, in which events, actions and agents combine in ways that are not always predictable or under control. Regrettably, accidents happen in spite of constant improvements in technical and administrative solutions. Achieving a comprehensive understanding of marine accidents and their development requires knowledge of the different elements involved in an accident event and the relationships between these elements. For example, to establish an overview of a collision accident, a description should include the cause of the event and its effects, as well as the entities involved.

In the field of maritime safety, recent research has focused on exploring the root causes of merchant shipping accidents. Studies have generally employed certain well-established mathematical models of risk assessment, such as Reason's (1990) Swiss Cheese model and the Bayesian Network model (cf. e. g. Pearl 2000), or frameworks, such as Human Factors Analysis and Classification System (HFACS) introduced by Shappell/Wiegmann (2000). Although these models have been widely implemented and despite the invaluable contributions in the field of safety science, some studies argue that the models have been interpreted in a linear way and the common understanding of an accident event is too simple (e. g. Schröder-Hinrichs/Hollnagel/Baldauf 2012: 156, 160). Also, researchers often focus on providing risk probabilities rather than on presenting background knowledge or defining causal concepts (cf. Mazaheri et al. 2015: 202, 204). Besides, the studies utilising these models and frameworks mainly rely on the intuition of the developers, or manual information extraction from accident reports, which is laborious and prone to subjectivity.

Unlike most of the earlier studies in the field of maritime safety, our study is corpus-based and lexicon-driven, which means that, instead of existing frameworks or classifications, we take a corpus as a starting point for our study and employ the FBT approach to compile predicate-argument structures, which manifest conceptual representations in the domain. Thus, the resulting conceptual representations are supposed to mirror real accident scenarios including concepts, conceptual relations, and the semantic roles of the entities involved. In this respect, the results of our study are intended to complement the existing domain ontologies and frameworks by providing a linguistically-based event representation. Also, instead of manual information extraction, we utilise a corpus tool and word sketches for the extraction of terminological information, i. e. information about concepts and their relations (cf. e. g. Leon-Aráuz/San Martín/Faber 2016). Word sketches are defined by Kilgarriff et al. (2014: 9) as summaries of a word's grammatical and collocational behaviour.

In the domain of maritime safety, an important source of domain-specific information are investigation reports. These reports are a rich source of data and include a meticulous analysis of individual accident cases. Investigation reports have previously been used as research material in other academic fields, such as the social sciences (e. g. Tang et al. 2013) and safety science (e. g. Mazaheri et al. 2015); however, it appears that they have not been analysed using terminological methods. Thus, a corpus study that includes an in-depth concept analysis should provide greater insight into the development of a marine accident.

The remaining sections of this paper are structured as follows: section 2 presents the the-

oretical background underlying the study and provides a short review of previous research and the applications of semantic frames in Terminology; section 3 illustrates the features of the corpus used in this study and describes the applied methodology; section 4 summarises and illustrates the results of the frame-based analysis; and section 5 presents the conclusions that can be derived from this research.

2 Frame-based Terminology

Frame-based Terminology is a cognitive approach to terminology that links specialised knowledge representation to Frame Semantics (cf. e. g. Fillmore 2006 [1982], Fillmore/Atkins 1992, Faber 2012). In Frame Semantics, frames are a structured way to present a scene or situation and are motivated by previous knowledge and experience. Another principle stemming from research shows that, regardless of the language, a set of verbs can be viewed as semantically related, as the verbs evoke the same general schematised scene (cf. Fillmore 2006 [1982]: 378, cf. also Croft/Cruse 2004: 8).

A number of studies in the field of FBT have demonstrated that the principles of Frame Semantics are applicable to specialised fields (cf. e. g. Faber/Márquez/Vega 2005, Pimentel 2014, Durán-Muñoz/L'Homme 2020). The earliest applications of FBT are the studies by Faber et al. (2006) in the domain of coastal engineering and by Faber (2012, 2014) in domain of the environment. EcoLexicon, a multilingual terminological knowledge base developed by the LexiCon Research Group at the University of Granada[1] (cf. e. g. Faber/Buendía Castro 2014), is likely to be the most extensive practical application of the methodology based on FBT. In addition to the work by the LexiCon Research Group, L'Homme (e. g. 2018) has applied a more FrameNet-oriented type of FBT in resources such as Framed DiCoEnviro. FrameNet[2] is a practical application of predicate frames originating from Fillmore's (1968) case frames and is often used as a tool in lexicography-oriented studies (cf. Fillmore/Johnson/Petruck 2003). Durán-Muñoz/L'Homme (2020) have also applied corpus-based analysis to English motion verbs in the field of adventure tourism. Researchers that have applied FBT in the field of law include Peruzzo (2014) and Pimentel (2014). However, in the specialised domains of transport, including aviation, railway, and maritime transport, the application of FBT is limited.

Events can be represented by predicate frames. Predicate frames stem from Fillmore's (1968) case frames in which verbs are characterised in terms of the semantic roles of their arguments (Faber/Reimerink 2019: 19). In FBT, the predicate frames characterising events, actions, and processes in a specialised domain are manifested by verbs and their nominalisations. Verbs also link the conceptual categories of the typical participants. The linguistic realisations of frame elements connect the linguistic level to the abstract conceptual representation of the situation (cf. Sánchez-Cárdenas/Ramisch 2019: 5). For example, in the field of maritime safety, a grounding event is an instance of a marine accident frame that includes cause, patient, and effect as core elements and symptoms and consequences as non-core elements (Pasanen 2014: 496, 498).

Our study focuses on predicate frames that describe actions and processes designated by verbs (cf. Faber/Cabezas-García 2019: 205). This starting point is based on the assumption that predicates and their arguments correspond to generic cognitive structures (Sán-

[1] http://lexicon.ugr.es/.

[2] https://framenet.icsi.berkeley.edu/fndrupal.

chez-Cárdenas/Ramisch 2019: 4). The arguments of the predicates include the frame elements that represent the main participants of a schematised situation. Moreover, the semantic nature of the arguments restricts the meaning of the predicate in the specialised domain (Faber/Cabezas-García 2019: 207). Linguistically, the arguments are generally nouns or noun phrases (Sánchez-Cárdenas/Ramisch 2019: 5).

3 Materials and methods

Within a specific domain, the specialised knowledge shared by domain experts is generally expressed in the texts produced by a community of these experts. For this reason, terminological analysis should be rooted in texts produced by domain experts (Peruzzo 2014: 152). Consequently, corpora are viewed as the main repositories of knowledge in a terminological project (L'Homme 2018: 8). A specialised corpus suitable for the purposes of this research was not available; therefore, the first step in our study was to compile a corpus.

In a terminological project, corpus texts should be selected according to criteria such as authoritative source, length, relationship with the topic, and textual genre. In the field of maritime safety, investigation reports are the main source of texts that meet these criteria and they are freely available for compiling a corpus on marine accident events. Consequently, the texts selected for the corpus in this study were all drawn from investigation reports and, therefore, all belong to the same textual genre. Investigation reports are structured documents that detail the findings of an investigation and thus represent a specialised communicative situation. The authors of these reports are investigation boards that are separately located in maritime countries, although currently many European investigation boards use the European Marine Casualty Information Platform (EMCIP) database of the European Maritime Safety Agency (EMSA) as a repository for the final reports.

The EMSA portal was used to compile the corpus for our study.[3] The filters that were activated included a casualty with a ship and cargo ship as the ship type. This search produced 296 results that covered the years 2012 to 2018 for all accident types. The reporting language is not included in the filtering options; therefore, the results were searched manually to identify the language of the investigation reports. Following the manual search, the reports provided by the Marine Accident Investigation Branch of the United Kingdom (MAIB) and the Maltese Safety Investigation Unit were selected, first, because of the high number of reports generated by these investigating bodies and, second, because all the reports were written in English.[4] Based on these criteria, 124 reports were selected for the corpus. The number of tokens was 877,693 and this total was considered sufficient when taking into account the specialised nature of the topic.

The documents downloaded from the EMCIP database have been systematically compiled by the investigation boards (Mazaheri et al. 2015: 209). A systematic working method means that all the reports presented by each investigation board have a uniform structure. This uniformity is due to the EMCIP taxonomy of input data. Introduced in 2011, the aim of the taxonomy is to facilitate the production of investigation reports and to unify their structure. By defining attributes and values that can be used to describe an occurrence, the taxonomy

[3] http://www.emsa.europa.eu/emcip.html.

[4] Cf. www.maib.gov.uk and https://mtip.gov.mt/en/Pages/MSIU/Marine-Safety-Investigation-Unit.aspx.

provides a certain standard for reporting and analysis.[5] However, Mazaheri (2017: 18) underlines the fact that accident reports are secondary sources that are prepared and collected using primary data. As secondary resources, they inevitably include a level of interpretation. Nevertheless, as Mazaheri (2017: 20) concludes, using secondary sources to analyse marine accidents is unavoidable as obtaining primary data is almost impossible.

The investigation reports published on the EMCIP portal are in PDF format. Therefore, this study first converted the reports into raw text. The documents were then loaded into the source code editor Notepad++.[6] In this application, the captions of the illustrations and figures, as well as the footnotes, were deleted unless they were in the form of complete sentences. Furthermore, texts in the illustrations, figures indicating pages or footnotes, and regular expressions, such as liability clauses, were removed. The chapters involving ship particulars, voyage particulars, and marine occurrence information, as well as contents, glossaries, annexes, and tables were also deleted.

In order to extract predicate-argument structures, the study's corpus was uploaded to the corpus query and management system Sketch Engine (Kilgarriff et al. 2015: 66; for a detailed introduction to the tool cf. Kilgarriff et al. 2014).[7] In the Sketch Engine corpus analysis application, the corpus was tagged with TreeTagger (version 2) as developed by Helmut Schmid with modifications by Sketch Engine.[8] Following the encoding using UTF-8 character, the corpus was then ready for querying and the generation of word sketches. First, we extracted a list of verbs from the corpus that included 1,122 lemmatised verbs. The 50 most frequent verbs were selected for the analysis. Table 1 shows the list of verbs and their frequency of occurrence with the topmost positions occupied by general language verbs, such as *be, have, do, take, make, use*. The general language verbs were included in the analysis because these polysemic verbs have potential special use in specific language when combined with specialised arguments (cf. Faber/Cabezas-García 2019: 208).

Table 1: 50 most frequent verbs in the corpus of Marine Accident Reports

No	Verb	Freq
1	*be*	32,483
2	*have*	9,006
3	*do*	1,618
4	*take*	1,569
5	*make*	1,279
6	*use*	1,267
7	*include*	1,159
8	*require*	1,154
9	*follow*	1,086

No	Verb	Freq
10	*provide*	1,006
11	*carry*	948
12	*indicate*	780
13	*report*	734
14	*identify*	688
15	*operate*	678
16	*ensure*	638
17	*give*	624
18	*show*	583

No	Verb	Freq
19	*remain*	563
20	*consider*	548
21	*fit*	532
22	*find*	532
23	*hold*	506
24	*leave*	500
25	*work*	495
26	*state*	484
27	*keep*	484

5 Cf. http://www.emsa.europa.eu/we-do/safety/accident-investigation/item/3024-emcip-taxonomy.html.

6 https://notepad-plus-plus.org/.

7 https://www.sketchengine.eu.

8 Cf. https://www.ims.uni-stuttgart.de/en/research/projects/textkorpora-werkzeuge/ and https://www.sketchengine.eu/english-treetagger-pipeline-2/.

No	Verb	Freq
28	*set*	483
29	*issue*	469
30	*pass*	452
31	*load*	415
32	*complete*	403
33	*see*	389
34	*conduct*	388
35	*start*	385

No	Verb	Freq
36	*result*	380
37	*call*	374
38	*inform*	369
39	*enter*	366
40	*maintain*	366
41	*go*	364
42	*monitor*	362
43	*check*	359

No	Verb	Freq
44	*confirm*	359
45	*cause*	353
46	*lead*	349
47	*avoid*	346
48	*reduce*	329
49	*become*	323
50	*increase*	318

In the next phase of the study, the word sketch function of the Sketch Engine corpus tool was used to search the corpus and find all the recurring patterns for each verb examined in the study (Kilgarriff et al. 2014: 10). The default word sketches represent the relations of verb-subject and verb-object pairs, as well as modifiers of the verb (Kilgarriff et al. 2014: 9, Faber/Cabezas-García 2019: 203). If the word being analysed is a verb, the left-hand column of the summary gives the objects of the verb (cf. Figure 1). Kilgarriff et al. (2014: 9) noted that "[t]he 'object' column is noise-free" and it was also identified useful for this study. Another effective word sketch is that of phrasal verbs (Kilgarriff et al. 2014: 10). The Sketch Engine tool also includes filters that assist with sorting the query results and focus on the most relevant output.

Figure 1: Word sketch of the verb carry *from Sketch Engine*

The next step in the search for predicate-argument structures was to analyse concordances of the most frequent verb-subject and verb-object pairs. We focused on the noun phrases that functioned as subjects and objects because the concordances that included these candidate terms would most likely produce valid predicate-argument structures. The threshold for candidate terms was set at four instances within the corpus. The results of concordance searches provided propositional representations of the processes; that is, the searches provided lexical representations of argument-taking lexical items, which are the verbs in our study, and their arguments. Knowledge of the types of entity that can fill the slots for the arguments (of the

predicates) is required for this process. After identifying potential arguments, we grouped them into broader conceptual categories as shown in Table 2, which presents the conceptual categories of potential human-related arguments (cf. Faber/Cabezas-García 2019: Table 2, Sánchez-Cárdenas/Ramisch 2019: Table 1). The conceptual categories were based on general semantic categories that include entities (animate and inanimate, concrete and abstract), states, processes, activities, events, and properties. The selected terms and their conceptual categories were later employed to assist the compilation and generalisation of argument structures (cf. section 4).

Table 2: Conceptual categories of potential human-related arguments

Category	**Subcategories**	**Potential arguments**
INDIVIDUAL HUMAN BEING(S)	on board	*chief engineer, chief officer, crew, crew member, engineer, lookout, master, mate, officer, pilot*
	not on board	*port captain, stevedore, surveyor*
ORGANISATION / COMPANY		*charterer, classification society, coastguard, flag state, harbour authority, owner, ship manager*
HUMAN STATE		*competence, fatigue, situational awareness*
HUMAN ACTIVITY	accident-related	*fire-fighting, investigation, rescue*
	cargo-related	*cargo operation, carriage, discharge, lashing, loading, sounding, stowage*
	navigation-related	*collision avoidance, navigation, navigational watch, manoeuvre, passage planning, pilotage, steering, watch-keeping*
	vessel-related	*audit, drill, inspection, maintenance, mooring, repair, risk assessment*
PRODUCT OF HUMAN ACTIVITY		*certificate of competency, loading manual, passage plan, recommendation, safety investigation report, safety management system, weather forecast*

The specification of argument structures involves identifying the semantic roles of the arguments based on the semantic relation between the argument and the predicate (Faber/Cabezas-García 2019: 202). Saeed (2016: 154) noted that semantic roles and grammatical relations have typical matchings; in a sentence, the subject often corresponds to the AGENT, the direct object to the THEME, and the INSTRUMENT is expressed by a prepositional phrase. Our classification of semantic roles is, in general outline, based on the VerbNet network of English verbs.[9] The semantic roles identified in our corpus are shown in Table 3.

[9] Semantic roles are dealt with in a number of sources; for example, in VerbNet: https://verbs.colorado.edu/verbnet.

Table 3: Semantic roles of the arguments (cf. Saeed 2016: 150–155, Faber/Cabezas-García 2019: 202–203, Sánchez Cárdenas/Ramisch 2019: 19)

Semantic role	Description
agent [actor]	a volitional participant, an entity, or a natural force, that initiates an action or event
beneficiary [recipient]	the entity for whose benefit the action was performed
experiencer	the living entity that experiences the action or event; is aware of the action or state; undergoes a sensory, cognitive, or emotional experience
force	non-volitional force, process, or event that produces a new entity or transforms a patient, affects a theme, or produces a result
instrument	an object or means used by an agent to perform an action or event
location	a place where an event occurs or where an object exists; path, source, goal
manner	the way in which an action takes place
patient	an entity affected by an event; undergoes transformation or change in state
purpose	the reason for which the action was performed
result	an event by an agent or a force; an entity originated or produced by an event

Table 4: Argument structures of the verb to carry

1ST ARGUMENT	PREDICATE	2ND ARGUMENT
[**artefact_vehicle**] *vessel, ship*	*carry* [**movement**]	[**product of human activity_cargo**] *cargo, general cargo, dangerous goods, forest products, Dangerous Chemicals in Bulk, vehicles, solid bulk cargo, temperature-controlled cargoes, dry cargo in bulk, heavy cargo, high density cargo, containers, coal in bulk, cargo of vegetable oil, U-IBA, cargo of steel turnings, heavy fuel oil, heavy and large volume cargoes, chemicals, products*
		[**equipment**] *radio equipment, spare full BA sets, deck lifter machine*
		[**device**] *Electronic Chart Display Information System (ECDIS), lifebuoy, lifeboat, immersion suit, firemen's outfit, AIS*
		[**product of human activity_document**] *loading manual, paper chart*
		[**human being**] *additional third officer, Filipino deckhand, watchkeeper, bosun, AB*
[**water movement**] *flood stream*	*carry* [**movement**]	[**artefact_vehicle**] *vessel, ship*

The annotation process based on the concordance searches allowed us to identify the arguments and show the relations between the predicates and the participants in the event. Table 4 presents the results of the annotation process of the verb *to carry*. In relation to an artefact,

namely a vessel or a ship in this context, the propositions with the verb *to carry* give information about the transported cargoes, the vessel's equipment and devices, as well as the crew members, passengers, or documents on board the vessel. The verb *to carry* is usually related to a vessel or a ship that has the purpose of carrying cargoes from port to port. However, the verb can also be related to a natural agent, such as a flood stream, that can carry vessels out of the fairway (cf. Table 4).

To demonstrate how linguistic information is connected to knowledge in the specialised field of maritime safety, the verb-argument structures belonging to the same lexical domain were grouped together. For the verb classification, we utilised general lexical domains that are widely used in studies on English verbs (Faber/Mairal 1999, Levin 1993, cf. also Faber/Cabezas-García 2019: 205, Faber/Reimerink 2019: 21). The lexical domains of FEELING, SOUND, and LIGHT mentioned by Faber/Mairal (1999) were excluded in our analysis due to the lack of verbs belonging to these domains in the list of 50 most frequent verbs in our corpus. Also, the lexical domain of CONTACT was excluded for the same reason. In our corpus, the domain of CONTACT is lexicalised in the combination of the support verb *make* and the noun *contact* as demonstrated in 4.1.6 and 4.1.8.

Each argument belongs to a specific conceptual category, has a semantic role, and is related to one or more concepts through a predicate; the predicate is a verb, a phrasal verb, or a verb pattern. For example, the argument *officer* belongs to the category of HUMAN BEING and has the semantic role of agent in the lexical domain of MANIPULATION; in the domain of MANIPULATION, the officer *operates safety equipment*; whereas the argument *safety equipment* belongs to the category of EQUIPMENT and has the role of instrument. In the following section, the results of the analysis are discussed and the evolving prototypical predicate-argument structures are drafted.

4 Results and discussion

4.1 The Marine Accident Event

The following section presents the results of the corpus search and includes the argument structures of a marine accident event. The argument structures are grouped under the main lexical domains of ACTION, CHANGE, EXISTENCE, MANIPULATION, MOVEMENT, PERCEPTION, POSSESSION, and SPEECH. As expected, some verbs belong to more than one lexical domain due to their multiple meanings; this is a particular feature of general language verbs (e. g. *take* and *make*). Moreover, some verbs are phrasal verbs, although in Table 1 they are listed in a base form. For example, these verbs include *carry* and its phrasal form *carry out*. Combinations of a general language verb (a support verb) and a noun, such as *take action, take place,* or *make contact,* are treated as verb patterns; in a text, they could be replaced by the verbs *act, happen,* or *contact*.

In the argument structures, the abbreviation *Arg* stands for the word *argument* and *As* stands for an *argument structure*. The second line in an argument structure is an instance from the corpus and is given to illustrate the structure.

4.1.1 Action

The lexical domain of ACTION belongs to the FrameNet frame of INTENTIONALLY_ACT that includes the core element of an agent performing an intentional act. In this analysis, performing an act includes starting and finishing the activity. In addition, verbs related to an ongoing activity (keep, maintain) belong to this domain.[10] In our corpus, the predicate verbs or phrasal verbs *carry out*, *conduct*, *complete*, *keep*, *maintain*, and *start* indicate a type of action that is taking place in the situation. The verb *take* is included in this domain as a support verb; this is because the main semantic frame is introduced by a noun (action) following the verb in the sentence (cf. Atkins/Fillmore/Johnson 2003: 270). In relation with a human agent, these verbs or phrasal verbs provide information about the actions of individual persons or groups of people, often in passive. Based on the concordances of these predicates, the following argument structures were developed:

(As 1 act)	TAKE ACTION (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{purpose}}$
	TAKE ACTION (bridge team)$_{\text{human agent}}$ (avoid/prevent an accident)$_{\text{purpose}}$
(As 2 act)	CARRY OUT (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{result}}$
	CARRY OUT (technician)$_{\text{human agent}}$ (maintenance)$_{\text{result}}$
(As 3 act)	COMPLETE (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{result}}$
	COMPLETE (officer)$_{\text{human agent}}$ (paperwork)$_{\text{result}}$
(As 4 act)	CONDUCT (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{result}}$
	CONDUCT (surveyor)$_{\text{human agent}}$ (audit)$_{\text{result}}$
(As 5 act)	KEEP (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{theme}}$
	KEEP (watchkeeper)$_{\text{human agent}}$ (navigational watch)$_{\text{theme}}$
(As 6 act)	MAINTAIN (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{theme}}$
	MAINTAIN (officer)$_{\text{human agent}}$ (lookout)$_{\text{theme}}$
(As 7 act)	START (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{theme}}$
	START (OOW)$_{\text{human agent}}$ (course alteration)$_{\text{theme}}$
	START (Arg 1)$_{\text{agent}}$ (Arg 2)$_{\text{theme}}$
	START (vessel)$_{\text{agent}}$ (a turn to port)$_{\text{theme}}$

In the first action scenario (As 1 act), a crew member or the crew collectively, a company or an organisation (Arg 1) takes an action to avoid or prevent an undesired event, such as a collision or an accident. This argument structure is normally followed by the applied measures. In the following scenarios (As 2–4 act), a human agent carries out, completes, or conducts an action that is part of the standard operation of the vessel, such as drills or maintenance; these are indicated with (Arg 2) in the argument structure. In these scenarios, the agent completes the activities. With the verb *conduct* as a predicate, the agent is often a company or an organisation rather than a crew member. Consequently, the action conducted is related to audits, inspections, and surveys (As 4 act). The predicates *keep* and *maintain* refer to ongoing navigation-related activities, particularly the action of keeping a lookout or watch. The navigation-related activities are ongoing and only cease once the navigation itself has ended (As 5–6 act). In our corpus, the verb *start* (As 7 act) is generally related to course alterations that are actions taken to prevent an accident. Thus, this verb can also indicate a change.

[10] https://framenet.icsi.berkeley.edu/fndrupal/framenet_search. All references to FrameNet refer to this source.

The conceptual categories of the arguments are interesting as a vessel or a ship, which belong to the conceptual category of VEHICLE, are often the grammatical subject in the argument structure of an action verb instead of a crew member or the crew (As 7 act). Therefore, a vessel or a ship represents the crew collectively and acts as an agent.

4.1.2 Change

In FrameNet, the frame of CHANGE has sub-frames CAUSE_CHANGE and UNDERGO_CHANGE in which an entity changes its category membership or the value of an attribute. The core elements are an entity and the initial and final categories or values. In the CAUSE_CHANGE frame, an agent is a core element. In navigation, the verbs *increase* and *reduce* often refer to the speed of the vessel (As 1–2 ch). Altering speed and course is the primary method for changing the movements of a vessel in a heavy traffic situation or in difficult weather conditions, thereby reducing the risk of an accident. In addition to the actions of crew members, natural phenomena may also affect navigation, such as fog reducing visibility.

(As 1 ch) INCREASE $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
INCREASE $(\text{master})_{\text{agent}}$ $(\text{speed})_{\text{theme}}$
(As 2 ch) REDUCE $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
REDUCE $(\text{master})_{\text{agent}}$ $(\text{speed})_{\text{theme}}$
REDUCE $(\text{fog})_{\text{agent}}$ $(\text{visibility})_{\text{theme}}$

The verb *have* is included in this domain as a support verb; this is the same application as the verb *take* in the ACTION domain (cf. 4.1.1). The main semantic frame of the verb *have* is introduced by a noun (effect, impact, influence) following the verb in the sentence. Man-made artefacts or constructions have a neutral, positive, or negative effect (impact, influence) on natural entities and can account for changes in the entity. However, the investigation reports often focus on the effect, impact, or influence of natural entities or events on other natural entities or water events (As 3 ch). In addition, artefacts or natural entities may influence water events by producing a change in the normal course of a process. Alternatively, a natural entity may have an effect on an artefact. During one event, for example, flood water on board the ship had a positive effect on the ship's stability.

(As 3 ch) HAVE EFFECT $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{patient}}$ OR $(\text{Arg 3})_{\text{theme}}$
HAVE EFFECT $(\text{dewatering system})_{\text{agent}}$ $(\text{beach})_{\text{patient}}$
HAVE IMPACT $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{patient}}$ OR $(\text{Arg 3})_{\text{theme}}$
HAVE IMPACT $(\text{wave})_{\text{agent}}$ $(\text{coastline})_{\text{patient}}$
HAVE INFLUENCE $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{patient}}$ OR $(\text{Arg 3})_{\text{theme}}$
HAVE INFLUENCE $(\text{nearshore sand bars})_{\text{agent}}$ $(\text{wave breaking})_{\text{theme}}$

Investigators have put considerable effort into studying the interactions between natural entities or events that may have had a role in the development of the accident (As 4 ch). In addition, human action or state may or may not have an effect (impact, influence) on human actions.

(As 4 ch) HAVE EFFECT $(\text{Arg 1})_{\text{force}}$ $(\text{Arg 2})_{\text{theme}}$
HAVE EFFECT $(\text{tidal velocities})_{\text{force}}$ $(\text{dune erosion})_{\text{theme}}$
HAVE EFFECT $(\text{fatigue})_{\text{force}}$ $(\text{decision making})_{\text{theme}}$
(As 5 ch) MAKE $(\text{Arg 1})_{\text{human agent}}$ $(\text{Arg 2})_{\text{result}}$
MAKE $(\text{bridge team})_{\text{human agent}}$ $(\text{course alteration})_{\text{result}}$

The verb *make* can be seen as a support verb when combined with the noun phrase *course alteration*; this combination could be replaced with the predicate verb *alter* and the noun argument *course* (As 5 ch).

4.1.3 Existence

In the field of maritime safety, the lexical domain of EXISTENCE encompasses incidents involving human actions and natural events that take place in a certain time or at a certain location, produce a result, and affect a patient. Investigators are interested in the communication between crew members and their situational awareness before, during, and after an accident event as failure to communicate effectively can *cause* or *lead to* misunderstandings (As 1–2 ex). The verb *to cause* is sometimes related to natural entities and events but also indicates the relation between the accident and an artefact-related event or a human action. Artefact-related events can include an electrical defect or fault, dust formation, cargo shift, refloating, and causing a damage or failure. Human actions can include erroneous actions of the captain, the chief engineer and the ship's crew, and a poor emergency response from the crew members. Human activities (heaving-to, lifting the stern ramp, ballasting) may cause undesired phenomena as the ship starts to drift, list, roll, vibrate, accelerate, or slide (As 1 ex). The phrasal verb *lead to* is often related to natural phenomena and the result is generally an artefact-related or natural event (As 2 ex).

(As 1 ex) CAUSE $(\text{Arg 1})_{\text{force}}$ $(\text{Arg 2})_{\text{result}}$ $(\text{Arg 3})_{\text{patient}}$ OR $(\text{Arg 3})_{\text{location}}$
CAUSE $(\text{fire})_{\text{force}}$ $(\text{damage})_{\text{result}}$ $(\text{engine room})_{\text{location}}$
CAUSE $(\text{Arg 1})_{\text{force}}$ $(\text{Arg 2})_{\text{result}}$
CAUSE $(\text{loss of situational awareness})_{\text{force}}$ $(\text{accident})_{\text{result}}$
CAUSE $(\text{ballasting})_{\text{force}}$ $(\text{listing})_{\text{result}}$

(As 2 ex) LEAD TO $(\text{Arg 1})_{\text{force}}$ $(\text{Arg 2})_{\text{result}}$
LEAD TO $(\text{lack of communication})_{\text{force}}$ $(\text{lack of mutual awareness})_{\text{result}}$
LEAD TO $(\text{slow leak})_{\text{force}}$ $(\text{build-up of pressure})_{\text{result}}$

The phrasal verb *result in* is commonly related to (erroneous) human actions (course alteration) or failures that result in losing control of the vessel or increase the risk of collision (As 3 ex).

(As 3 ex) RESULT IN $(\text{Arg 1})_{\text{force}}$ $(\text{Arg 2})_{\text{result}}$
RESULT IN $(\text{accident})_{\text{force}}$ $(\text{spill})_{\text{result}}$
RESULT IN $(\text{operational failure})_{\text{force}}$ $(\text{risk of collision})_{\text{result}}$
RESULT IN $(\text{Arg 1})_{\text{force}}$ $(\text{Arg 2})_{\text{result}}$ $(\text{Arg 3})_{\text{patient}}$
RESULT IN $(\text{water ingress})_{\text{force}}$ $(\text{damage})_{\text{result}}$ $(\text{vessel})_{\text{patient}}$

In addition to the human actions on board the vessel before and after the accident, investigators thoroughly analyse natural events that take place in the accident location or in other locations that resemble the accident position. The natural events mentioned are often related to erosion or waves (As 4 ex). The phrase *take place* is a combination of the support verb *take* and the noun *place* and has the basic meaning 'to happen'. Therefore, we treat the phrase as one lexical unit.

(As 4 ex) TAKE PLACE (Arg 1)$_{theme}$ (Arg 2)$_{time}$
TAKE PLACE (inspection)$_{theme}$ (annually)$_{time}$
TAKE PLACE (Arg 1)$_{theme}$ (Arg 2)$_{location}$
TAKE PLACE (wave breaking)$_{theme}$ (over the shoals)$_{location}$
(As 5 ex) ENSURE (Arg 1)$_{agent}$ (Arg 2)$_{theme}$
ENSURE (crew)$_{agent}$ (operation of a vessel)$_{theme}$
ENSURE (VTS)$_{agent}$ (safety of navigation)$_{theme}$
(As 6 ex) SHOW (Arg 1)$_{agent}$ (Arg 2)$_{theme}$
SHOW (vessel)$_{agent}$ (navigation lights)$_{theme}$

In the marine transportation process, the crew (Arg 1) ensures safe and professional operation of the vessel (Arg 2). In addition, human actions can be given the role of an instrument to ensure the existence of a certain state or situation, such as safe operation of the ship and the activities on board. Services that have a purpose to ensure safety of navigation (Arg 2) are also present and two examples of these are marine forecast systems and vessel traffic services (VTS) (Arg 1). The domain of EXISTENCE also encompasses verbs that indicate the perception of an entity (cf. Faber/Mairal 1999: 279). The verb *show* indicates the existence of the navigation lights for a potential observer in (As 6 ex).

4.1.4 Manipulation

The lexical domain of MANIPULATION contains various argument structures that account for the different dimensions of a vessel's operation. This lexical domain also allows for human agents (Arg 1) on board vessels or ashore. In the former case, the agent is usually the master or a crew member who operates the vessel or the equipment (Arg 2) on board in a specific manner (Arg 3). In the latter case, the agent is a company that operates vessels in a defined area and in a specific manner (Arg 3); for example, worldwide, commercially, under the Maltese flag, on international trade, on the spot market, under charter, under the company's safety management system (SMS), or in compliance with stated safety requirements. The instrument can take the grammatical place of the actual agent in a sentence, such as "the vessel operates worldwide". According to FrameNet, location, manner, and purpose are non-core elements in the frame of MANIPULATION. However, in our corpus the complements *in restricted waters* and *worldwide* in (As 1 man) are obligatory elements in the predicate-argument structures, since without the complements the sentences would be insignificant.

(As 1 man) OPERATE (Arg 1)$_{agent}$ (Arg 2)$_{instrument}$ (Arg 3)$_{location}$ OR (Arg 3)$_{manner}$
OPERATE (master)$_{agent}$ (vessel)$_{instrument}$ (in restricted waters)$_{location}$
OPERATE (Ø)$_{agent}$ (vessel)$_{instrument}$ (world-wide)$_{manner}$
(As 2 man) USE (Arg 1)$_{agent}$ (Arg 2)$_{instrument}$ (Arg 3)$_{purpose}$
USE (pilot)$_{agent}$ (radio)$_{instrument}$ (inform the coastguard)$_{purpose}$
USE (vessel)$_{agent}$ (traffic separation scheme)$_{instrument}$
USE (OOW)$_{agent}$ (radar)$_{instrument}$ (determine the risk of collision)$_{purpose}$

In the argument structure (As 2 man), the agent (Arg 1) is a crew member who uses an instrument (Arg 2) for a specific purpose (Arg 3). The instrument is usually a navigational device or a VHF radio, which is used to monitor traffic, to clarify the identity or intentions of another vessel, or to escape a hazardous situation such as manœuvring clear of the shallows. From the viewpoint of the investigation, the purpose of the action is a core element.

The actions of the crew members often include a change to the settings of the devices or equipment on board. These types of actions have the verb *to set* as the predicate in the argument structure (As 3 man). For example, the master or the officer may set the radar display north-up, the radio on a certain channel, or the propeller pitch to zero. If the device or equipment has been in the off position and is then started, the predicate verb *to start* indicates a change in the status of the device (As 4 man).

(As 3 man) SET (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{theme}}$ (Arg 3)$_{\text{result}}$
SET (master)$_{\text{human agent}}$ (engine)$_{\text{theme}}$ (to full astern)$_{\text{result}}$
(As 4 man) START (Arg 1)$_{\text{human agent}}$ (Arg 2)$_{\text{theme}}$
START (engineer)$_{\text{human agent}}$ (engine)$_{\text{theme}}$

The verb *to start* usually appears in a passive form and the indication that a human agent performs the action is only implicit (Ex 2).

4.1.5 Movement

In the prototypical MOTION frame, an entity with the role of a theme moves along a path from a source to a goal. Alternatively, the movement may take place within an area (cf. FrameNet s.v. *motion*). A vessel has a function to operate in specific sea areas with the aim to make her way at a defined speed and carry cargoes from one port to another. In this domain, the support verb *make* takes the noun *way* to introduce the main semantic frame (cf. Atkins/Fillmore/Johnson 2003: 270). While undertaking transport operations, vessels leave a port, pass waypoints, enter another port, or call at a port. Thus, vessels participate in the lexical domain of MOVEMENT. The argument structure involves an artefact in the role of a theme, as well as an argument indicating a source, path, or goal. In the verb-argument structures below, we have used the upper category of location rather than a source, path, or goal. Instead of a concrete geographical location, the location of an artefact may be artificial, such as a navigation track or a traffic lane. The argument structure may include non-core elements, such as a manner, purpose, or time.

As well as the argument structure that involves the core frame elements of an artefact and a location and the non-core elements of time and manner, the lexical domain of MOVEMENT contains human beings. Human beings make their way or go to another location on board a vessel, or enter or leave the location. Based on the concordance searches, the following argument structures evolved:

(As 1 mov) MAKE * WAY (Arg 1)$_{\text{theme}}$ (Arg 2)$_{\text{location}}$
MAKE * WAY (vessel)$_{\text{theme}}$ (into port)$_{\text{location}}$
MAKE * WAY (pilot)$_{\text{theme}}$ (to the bridge)$_{\text{location}}$
(As 2 mov) ENTER (Arg 1)$_{\text{theme}}$ (Arg 2)$_{\text{location}}$ (Arg 3)$_{\text{time}}$ (Arg 4)$_{\text{manner}}$
ENTER (vessel)$_{\text{theme}}$ (port)$_{\text{location}}$ (at 0640)$_{\text{time}}$ (safely)$_{\text{manner}}$
ENTER (Arg 1)$_{\text{theme}}$ (Arg 2)$_{\text{location}}$ (Arg 3)$_{\text{purpose}}$
ENTER (second engineer)$_{\text{theme}}$ (control room)$_{\text{location}}$ (to lock the propeller shaft)$_{\text{purpose}}$
(As 3 mov) GO (Arg 1)$_{\text{theme}}$ (Arg 2)$_{\text{location}}$
GO (second officer)$_{\text{theme}}$ (chart room)$_{\text{location}}$
(As 4 mov) LEAVE (Arg 1)$_{\text{theme}}$ (Arg 2)$_{\text{location}}$ (Arg 3)$_{\text{time}}$
LEAVE (lookout)$_{\text{theme}}$ (bridge)$_{\text{location}}$ (at 1013)$_{\text{time}}$

(As 5 mov) PASS (Arg 1)$_{theme}$ (Arg 2)$_{location}$ (Arg 3)$_{time}$
PASS (vessel)$_{theme}$ (waypoint)$_{location}$ (two hours after an event)$_{time}$
(As 6 mov) CALL AT (Arg 1)$_{theme}$ (Arg 2)$_{location}$
CALL AT (vessel)$_{theme}$ (port)$_{location}$

In the context of marine transportation, only vessels pass another vessel, area, or solid object (As 5 mov), or move, drift, or swing. Accordingly, only humans can go to another location (As 3 mov). In this context, the vessel is normally the means of conveyance of the theme, thereby fulfilling the role of a carrier. In addition to the vessel, a natural force may have the role of a carrier (As 7 mov).

(As 7 mov) CARRY (Arg 1)$_{agent}$ (Arg 2)$_{theme}$ (Arg 3)$_{location}$
CARRY (vessel)$_{agent}$ (general cargo)$_{theme}$ (to Northern Europe)$_{location}$
CARRY (flood stream)$_{agent}$ (vessel)$_{theme}$ (out of the fairway)$_{location}$
CARRY (vessel)$_{agent}$ (additional third officer)$_{theme}$

In the first scenario (As 7 mov), a vessel (Arg 1) carries cargo (Arg 2) from one geographical location to another; for example, from South America to Northern Europe (Arg 3). In the second scenario, a natural force (Arg 1) carries a vessel (Arg 2) in a certain direction (Arg 3). In the third scenario, a vessel (Arg 1) carries certain crew members on board, such as an additional third officer, a Filipino deckhand, a watchkeeper, a bosun, or an AB (Arg 2).

The verb *load* has been included in the lexical domain of MOVEMENT as the verb belongs to the FrameNet frames of PLACING, in which an agent places a theme at a location, and FILLING, in which a theme moves to a container. In FrameNet, both frames are related to the CAUSE_MOTION frame (FrameNet s.v. *load*).

(As 1 psn) LOAD (Arg 1)$_{agent}$ (Arg 2)$_{theme}$ (Arg 3)$_{location}$ (Arg 4)$_{manner}$
LOAD (vessel)$_{agent}$ (grain)$_{theme}$ (in Southampton)$_{location}$
LOAD (stevedores)$_{agent}$ (solid bulk)$_{theme}$ (in homogenous distribution)$_{manner}$

The argument structure with the verb *load* as the predicate (As 1 psn) may have either a vessel or a group of humans as the first argument (Arg 1) and in the role of an agent. The second argument (Arg 2) is the cargo that is loaded and it has the role of a theme. Interestingly, the first argument (Arg 1) limits the choice of the other arguments. The argument indicating location (Arg 3) is related to a non-human agent, which is a vessel in this context, and the argument indicating manner (Arg 4) is related to human agents.

4.1.6 Perception

In the FrameNet frame of PERCEPTION, a perceiver perceives a phenomenon. The perception may be passive or active (FrameNet s.v. *perception*). In our corpus, the active role of the perceiver is underlined by the use of the phrase *make visual contact with* (As 2 per) instead of the verb *see*. In the maritime safety context, the lexical domain of PERCEPTION contains the frame elements of a human being in the role of an agent, a water vehicle to be identified, a device to be monitored, or a natural entity. A human being (master, pilot, or officer) actively monitors or does not monitor on board devices or the surrounding navigational situation (As 4 per). A human being also checks or does not check the equipment settings, the status of equipment, the intended route of the vessel, the vessel's position, speed or movements, and natural or weather conditions in the location (As 5 per). Additionally, a human being identifies or finds defects,

deficiencies, or damages. In the engine department, the chief engineer has to check the status of the machinery and tanks, as well as steering and propulsion systems (As 5 per). The object of perception may also be damage that the artefact has suffered. The location of the object is sometimes included in the argument structure (As 1 per).

(As 1 per)	SEE $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$ $(\text{Arg 3})_{\text{location}}$
	SEE $(\text{crew member})_{\text{agent}}$ $(\text{smoke})_{\text{theme}}$ $(\text{in a cabin})_{\text{location}}$
(As 2 per)	MAKE VISUAL CONTACT WITH $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	MAKE VISUAL CONTACT WITH $(\text{pilot})_{\text{agent}}$ $(\text{fishing boat})_{\text{theme}}$
(As 3 per)	FIND $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	FIND $(\text{engineer})_{\text{agent}}$ $(\text{defect})_{\text{theme}}$
(As 4 per)	MONITOR $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	MONITOR $(\text{bridge crew})_{\text{agent}}$ $(\text{pitch indicator})_{\text{theme}}$
(As 5 per)	CHECK $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	CHECK $(\text{navigational officer})_{\text{agent}}$ $(\text{speed})_{\text{theme}}$
	CHECK $(\text{chief engineer})_{\text{agent}}$ $(\text{engine oil level})_{\text{theme}}$

The lexical domain of PERCEPTION is closely related to the domain of COGNITION that includes the use of one's mind for forming an idea of something. Perception is also related to inspections or audits, during which non-conformities of deficiencies are potentially identified. Therefore, assigning a verb to a single domain is not always a straightforward process. In this study, the verbs *monitor* and *check* have been included in the domain of PERCEPTION as visual monitoring of the devices is crucial in the navigational context, and the possible decisions to act are based on these perceptions.

4.1.7 Possession

In our corpus, the instances belonging to the lexical domain of POSSESSION have an argument structure that includes a human or an artificial agent and a theme. Examples of a theme include a device, product of human activity, information on a traffic situation, natural conditions, and water movements. These argument structures have the phrasal verb *fit with* or the verbs *carry*, *have*, *hold*, or *provide* as a predicate. In the first and second scenarios (As 1–2 poss), a vessel (bridge, engine) is fitted with, has, or carries certain devices, equipment, or documents, such as radio equipment, lifeboats, loading manual, paper charts, or immersion suits. In the third scenario (As 3 poss), a crew member (Arg 1) holds a type of certificate. The agent may also be equipment, or a man-made system or entity that includes certain parts (As 4 poss).

(As 1 poss)	FIT WITH $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	FIT WITH $(\text{vessel})_{\text{agent}}$ $(\text{double bottom})_{\text{theme}}$
(As 2 poss)	CARRY $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	CARRY $(\text{ship})_{\text{agent}}$ $(\text{certification})_{\text{theme}}$
(As 3 poss)	HOLD $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	HOLD $(\text{chief mate})_{\text{agent}}$ $(\text{endorsement})_{\text{theme}}$
(As 4 poss)	INCLUDE $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	INCLUDE $(\text{equipment})_{\text{agent}}$ $(\text{radar})_{\text{theme}}$
(As 5 poss)	PROVIDE $(\text{Arg 1})_{\text{agent}}$ $(\text{Arg 2})_{\text{theme}}$
	PROVIDE $(\text{London VTS's principal control centre})_{\text{agent}}$ $(\text{traffic information})_{\text{theme}}$

PROVIDE (wave buoy)$_{agent}$ (information)$_{theme}$
(As 6 poss) ISSUE (Arg 1)$_{agent}$ (Arg 2)$_{theme}$
ISSUE (company)$_{agent}$ (safety bulletin)$_{theme}$

In the argument structure (As 5 poss), the companies and organisations in the maritime industry (Arg 1) provide guidance and information on a certain topic, such as safe navigation, traffic situations, and meteorological situations (Arg 2). Navigational devices (Arg 1) provide data and information on traffic situations or weather conditions (Arg 2). Administrations, companies, agencies, or societies issue certificates, bulletins, circulars, and notices (As 6 poss).

4.1.8 Speech

The lexical domain of SPEECH represents scenarios that include a communicator or speaker directing communication containing a message to an addressee (see FrameNet frames CONTACTING and TELLING). In our corpus, the domain of SPEECH includes a scenario in which a person (Arg 1) on board the vessel makes a voice contact or radio contact with another ship or a coast station (Arg 2). Other scenarios include a person who reports a problem, situation, intent, or change of conditions to a receiver that may be another vessel or a shore-based station (Arg 2). The instrument used for the contact is sometimes indicated (Arg 3). In the second scenario (As 2 sp), a crew member, coast station, or shore-based service provider (Arg 1) calls another crew member, vessel, or shore-based station (Arg 2) using a radio (Arg 3). In the third scenario (As 3 sp), the intention of the agent (Arg 1) is to have the other person (Arg 2) move to another location (Arg 3). An instrument (Arg 4) may be needed if the person called is located a significant distance from the agent.

(As 1 sp) MAKE CONTACT WITH (Arg 1)$_{agent}$ (Arg 2)$_{experiencer}$ (Arg 3)$_{instrument}$
MAKE CONTACT WITH (pilot)$_{agent}$ (VTS)$_{experiencer}$ (mobile phone)$_{instrument}$
(As 2 sp) CALL (Arg 1)$_{agent}$ (Arg 2)$_{experiencer}$ (Arg 3)$_{instrument}$
CALL (VTS)$_{agent}$ (vessel)$_{experiencer}$ (on VHF channel 09)$_{instrument}$
(As 3 sp) CALL (Arg 1)$_{agent}$ (Arg 2)$_{theme}$ (Arg 3)$_{location}$ (Arg 4)$_{instrument}$
CALL (watchkeeper)$_{agent}$ (master)$_{theme}$ (to the bridge)$_{location}$ (by radio)$_{instrument}$
(As 4 sp) INFORM (Arg 1)$_{agent}$ (Arg 2)$_{beneficiary}$ (Arg 3)$_{theme}$
INFORM (master)$_{agent}$ (coastguard)$_{beneficiary}$ (collision)$_{theme}$
(As 5 sp) CONFIRM (Arg 1)$_{agent}$ (Arg 2)$_{theme}$
CONFIRM (OOW)$_{agent}$ (the change of the course)$_{theme}$

In the scenario (As 4 sp), a crew member, pilot, or another human agent (Arg 1) informs or does not inform the master, pilot, company, or shore-based service provider about, for example, a failure, an alarm, or an accident (Arg 3). In the scenario (As 5 sp), a human agent confirms by speech that a certain human activity has taken place.

4.2 Argument structure generalisation

Based on the concordance searches, it can be concluded that certain predicates are activated more frequently than others in the context of maritime safety (cf. Faber et al. 2005: 4). The marine accident frame includes the following: verbs such as *provide*, *ensure*, *indicate*, *require*, *operate*, *use*, *identify*, *report*, and *carry*; phrasal verbs such as *carry out*; and, verb patterns such

as *have an effect* and *make a contact*. These verbs, phrasal verbs, and verb patterns connect the arguments belonging to conceptual categories such as HUMAN BEINGS, HUMAN ACTIVITIES, ATMOSPHERIC PHENOMENA, WATER MOVEMENTS, EQUIPMENT, DEVICES, and VEHICLES, and they have the roles of agent, patient, theme, experiencer, location, force, purpose, result, instrument, beneficiary, or manner.

The marine accident event is conceptualised as a process that involves human agents, artefacts, and natural agents. Human agents start, carry out, and complete actions, take action to avoid events, operate and use instruments, identify and ensure a state, keep or maintain a watch, increase or reduce speed, and go from one place to another. Artefacts carry cargoes from one place to another, make a way or a contact, indicate or ensure a state, have an effect, an impact, or an influence, pass a waypoint, call at a port, and remain in a position. Natural agents take place or have an effect, an impact, or an influence, and cause or lead to a result.

In summary, a marine accident event includes a variety of participants, actions, and events, some of which lead to marine accidents while others focus on avoiding the accidents. The argument-predicate structures in the lexical domain of EXISTENCE illustrate the connection between the argument-predicate structures and a marine accident. The causal nature of marine accident events is shown through the verb *cause* and the phrasal verbs *lead to* and *result in* as they indicate the relation between a human action or an artefact-related event and the accident. These events are dynamic processes based on interacting components, the relations of which are constantly changing (cf. Faber et al. 2005). A natural phenomenon occurs and changes independently of human action. At the same time, human state changes, or a human takes action, and an artefact indicates, moves, and operates. These processes can affect the environment, humans, and artefacts. Effects can include the loss of a ship, a casualty with ships or cargo, pollution, fatalities, and injuries.

5 Conclusions

The aim of this study was to identify predicate-argument structures in the specialised discourse of maritime safety by using principles of Frame-based Terminology (FBT). The predicate-argument structures identified can be used to enhance the representation and understanding of marine accident events and potentially create a domain-specific frame for marine accidents. Thus, the predicate-argument structures are presented for consultation to future end users of the results, including the stakeholders in the field of maritime safety and terminologists.

Based on lexicon-driven analysis of a corpus consisting of marine accident investigation reports, we demonstrate that FBT can be successfully applied to a marine accident event. The methodology of information extraction by means of word sketches and corpus-based analysis has led to compilation of initial predicate-argument structures for a marine accident event. It was found that in this context, certain predicates are activated more frequently than others. A large number of these predicates are general language verbs that construct special meanings in combination with specialised arguments.

Avoiding future accidents is the common goal of all the stakeholders in the maritime field. Accident investigation is the primary method for gaining insight into how accidents develop. The corpus-based lexicon-driven analysis of a marine accident event, as manifested in accident investigation reports, shows that the chain of actions and events leading to an accident is complicated and involves a variety of elements. These elements can be natural and technical as well as human. Specifically, in the analysed accident investigation reports, it was found that a

vessel or a ship often represents the crew collectively and acts as the agent instead of the crew or crew members.

Furthermore, the relation between the marine accident event and a counteracting cause is important in the field of maritime safety as it helps researchers locate the measures that could have prevented the situation from developing and ending in an accident. Mazaheri et al. (2015: 206) call these measures safety factors that act as barriers and stop an incident from turning into an accident.

The environment in which a maritime transport event takes place can be divided into the following components: vessel, sea, and shore. The vessel and her interactions with the environment, maritime traffic, and the shore comprise a complex socio-technical system that is affected by environmental factors. It is intended that the results of this study will facilitate a better understanding of these complex systems and the processes that lead to marine accidents.

References

Atkins, Sue / Fillmore, Charles J. / Johnson, Christopher R. (2003): "Lexicographic Relevance: Selecting Information from Corpus Evidence." *International Journal of Lexicography* 16.3: 251–280.

Croft, William / Cruse, Alan (2004): *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Durán-Muñoz, Isabel / L'Homme, Marie-Claude (2020): "Diving into English Motion Verbs from a Lexico-semantic Approach. A Corpus-based Analysis of Adventure Tourism." *Terminology* 26.1: 33–59. DOI 10.1075/term.00041.dur.

EMSA [European Maritime Safety Agency] (n. d.): "An Agency of the European Union". http://www.emsa.europa.eu/ (27.12.2018).

Faber, Pamela (2014): "Frames as a Framework for Terminology." *Handbook of Terminology*. Eds. Frieda Steurs / Hendrik Kockaert. Amsterdam: Benjamins. 14–33.

Faber, Pamela (2018): "Frames and Specialized Knowledge Representation." *Terminologie & Ontologie: Théories et Applications. Actes de la conference TOTh 2018. Le Bourget du Lac 7–8 juin 2018.* Chambéry: Presses Universitaires Savoie Mont Blanc. 15–31.

Faber, Pamela, ed. (2012): *A Cognitive Linguistics View of Terminology and Specialized Language.* Berlin / New York: Mouton de Gruyter.

Faber, Pamela / Buendía Castro, Miriam (2014): "EcoLexicon." *Proceedings of the XVI Euralex International Congress: The User in Focus*. *15–19 July 2014, Bolzano/Bozen.* Eds. Andrea Abel / Chiara Vettori / Natascia Ralli. Bolzano: Euralex. 601–607. http://euralex2014.eurac.edu/en/callforpapers/Documents/EURALEX%202014_gesamt.pdf (05.03.2020).

Faber, Pamela / Cabezas García, Melania (2019): "Specialized Knowledge Representation: From Terms to Frames." *Research in Language* 17.2: 197–212. https://doi.org/10.2478/rela-2019-0012 (12.05.2021).

Faber, Pamela / Mairal Usón, Ricardo (1999): *Constructing a Lexicon of English Verbs.* Berlin: Mouton de Gruyter.

Faber, Pamela / Márquez Linares, Carlos / Vega Expósito, Miguel (2005): "Framing Terminology: A Process-oriented Approach." *Meta* 50.4. https://doi.org/10.7202/019916ar (25.8.2022).

Faber, Pamela / Montero Martínez, Silvia / Castro Prieto, María Rosa / Senso Ruiz, José / Prieto Velasco, Juan Antonio / León Arauz, Pilar / Márquez Linares, Carlos / Vega Expósito, Miguel (2006): "Process-oriented Terminology Management in the Domain of Coastal Engineering." *Terminology* 12.2: 189–213.

Faber, Pamela / Reimerink, Arianne (2019): "Framing Terminology in Legal Translation." *International Journal of Legal Discourse* 4.1: 15–46. https://doi.org/10.1515/ijld-2019-2015 (12.05.2021).

Fillmore, Charles J. (1968): "The Case for Case." *Universals in Linguistic Theory.* Eds. Emmon Bach / Robert Harms. New York: Holt, Reinhart & Winston. 1–81.

Fillmore, Charles J. (1977): "Scenes-and-frames Semantics." *Linguistics Structures Processing.* Ed. Antonio Zampolli. Amsterdam / New York: North Holland. 55–81.

Fillmore, Charles J. (1985): "Frames and the Semantics of Understanding." *Quaderni di Semántica* 6.2: 222–254.

Fillmore, Charles J. (2006) [1982]. "Frame Semantics." *Cognitive Linguistics: Basic Readings*. Ed. Dirk Geeraerts. Berlin: De Gruyter Mouton. 373–400. [Originally published in 1982 in: *Linguistics in the Morning Calm.* Ed. Linguistic Society of Korea. Seoul: Hanshin. 111–137.]

Fillmore, Charles / Atkins, Sue (1992): "Towards a Frame-based Lexicon: The Semantics of RISK and its Neighbors." *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organisation.* Eds. Adrienne Lehrer / Eva F. Kittay. Hillsdale, NJ: Erlbaum. 75–102.

Fillmore, Charles / Johnson, Christopher / Petruck, Miriam (2003): "Background to FrameNet." *International Journal of Lexicography* 16.3: 235–250.

FrameNet. https://framenet.icsi.berkeley.edu/fndrupal/ (25.10.2020).

Kilgarriff, Adam / Baisa, Vit / Bušta, Jan / Jakubiček, Miloš / Kovář, Vojtěch / Michelfeit, Jan / Rychlý, Pavel / Suchomel, Vit (2014): "The Sketch Engine: Ten Years on." *Lexicography ASIALEX* 2014.1: 7–36. DOI 10.1007/s40607-014-0009-9.

Kilgarriff, Adam / Marcowitz, Fredrik / Smith, Simon / Thomas, James (2015): "Corpora and Language Learning with the Sketch Engine and Skell." *Revue française de linguistique appliquée* XX.1: 61–80. https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2015-1-page-61.htm (04.02.2019).

León-Araúz, Pilar / San Martín, Antonio / Faber, Pamela (2016): "Pattern-based Word Sketches for the Extraction of Semantic Relations." *Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016).* Osaka: Coling. 73–82.

Levin, Beth (1993): *English Verb Classes and Alterations: A Preliminary Investigation.* Chicago: The University of Chicago Press.

L'Homme, Marie-Claude (2018): "Maintaining the Balance between Knowledge and the Lexicon in Terminology: A Methodology based on Frame Semantics." *Lexicography ASIALEX* 4: 3–21. https://doi.org/10.1007/s40607-018-0034-1 (12.05.2021).

Mazaheri, Arsham (2017): *A Framework for Evidence-based Risk Modeling of Ship Grounding.* (Aalto University publication series Doctoral dissertations 112/2017). Espoo: Aalto University. http://urn.fi/URN:ISBN:978-952-60-7478-8 (17.03.2019).

Mazaheri, Arsham / Montewka, Jakub / Nisula, Jari / Kujala, Pentti (2015): "Usability of Accident and Incident Reports for Evidence-based Risk Modeling – A Case Study on Ship Grounding Reports." *Safety Science* 76: 202–214.

Pasanen, Päivi (2014): "Knowledge Patterns as Indicators of Cause-and-effect Relations in the Domain of Maritime Safety." *Languages for Special Purposes in a Multilingual, Transcultural World. Proceedings of the 19th European Symposium on Languages for Special Purposes, 8–10 July 2013, Vienna, Austria.* Eds. Gerhard Budin / Vesna Lušicky. Vienna: University of Vienna. 490–499.

Pearl, Judea (2000): *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Peruzzo, Katia (2014): "Term Extraction and Management Based on Event Templates. An Empirical Study on an EU Corpus." *Terminology* 20.2: 151–170.

Pimentel, Janine (2014): "Using Frame Semantics to Build a Bilingual Lexical Resource on Legal Terminology." *Handbook of Terminology.* Eds. Frieda Steurs / Hendrik J. Kockaert. Amsterdam: Benjamins. 427–450.

Reason, James (1990): *Human Error*. Cambridge: Cambridge University Press.

Saeed, John (2016): *Semantics.* 4th ed. Chichester: Wiley Blackwell.

Sánchez-Cárdenas, Beatriz / Ramisch, Carlos (2019): "Eliciting Specialized Frames from Corpora Using Argument-structure Extraction Techniques." *Terminology* 25.1: 1–31.

Schröder-Hinrichs, Jens-Uwe / Hollnagel, Erik / Baldauf, Michael (2012): "From Titanic to Costa Concordia – a century of lessons not learned." *WMU Journal of Maritime Affairs* 11.2: 151–167.

Shappell, Scott A. / Wiegmann, Douglas A. (2000): *The Human Factors Analysis and Classification System–HFACS. Final Report DOT/FAA/AM-00/7.* Washington: Office of Aviation Medicine, Federal Aviation Administration.

Tang, Lijun / Acejo, Iris / Ellis, Neil / Turgo, Nelson / Sampson, Helen (2013): "Behind the Headlines? An Analysis of Accident Investigation Reports." *Seafarers International Research Centre Symposium Proceedings (2013).* Cardiff: Seafarers International Research Centre (SIRC), Cardiff University. 1–21.

Päivi Pasanen, PhD
University of Helsinki
Puhakantie 17
49490 Neuvoton
paivi.pasanen@helsinki.fi

Form – Meaning – Usage Synergy in LSP & Professional Communication: Computer Security in Terms of Frame Semantics

Ekaterina Isaeva, Olga Baiburova & Oksana Manzhula

Abstract This article deals with computer security terminology from the perspective of Fillmore's frame semantics. Cognitive linguists have proved that semantics is realised in context and contributes to cognition. This article aims to analyse which semantic roles are prevalent for using computer security terms in context. Our evidence proves that the term's meaning and semantic role in the frame corresponding to a communicative event are interrelated. The research has been carried out on a manually collected corpus of computer security texts, comprising terms in their contexts. The data are analysed as follows. First, a thematic categorisation of terms is carried out. Then semantic frame modelling is applied. After that, we generalise our findings and achieve sufficient abstraction in the conclusion about the presence of form – meaning – usage interdependence in professional discourse and LSP. Finally, we discuss the place of semantic framing in the multimodality of professional communication regarding the logic and philosophy of language.

Keywords computer security discourse, LSP, professional communication, semantic framing, semantic role, transdiscursive communication

1 Introduction

We present a study of computer security (CS) terminology, carried out within a project "Special Knowledge Mediation by Means of Automated Ontological and Metaphorical Modelling". The project aims to find effective ways to transfer, receive, process, and store specialised knowledge. This knowledge is acquired through professional experience, stored as mental models, and represented via a language for specific purposes (LSP). All these are integral components of professional communication, and each of them is crucial for successful knowledge transfer and acquisition.

This article examines form – meaning – usage interdependence in professional discourse. As we showed earlier (Isaeva 2019: 81), accidental cognitive framing of specialised concepts in transdiscursive (between experts and nonexperts) professional communication can cause significant loss of transferred information. Therefore, finding interdependence between grammar, semantics, and pragmatics, referred to as form, content, and usage, respectively, will be helpful in deliberate cognitive framing for enhancing professional communication. Additionally, the findings can be applied for text-mining and machine learning as far-sighted goals. Thus, the results, which tackle different aspects of linguistics, applied to the specific field of professional communication, i. e., CS, could be of interest to those involved in the studies of structural and derivational grammar, semantics, syntax, discourse, cognitive and computational linguistics, terminology, and mediation.

Zitiervorschlag / Citation:
Isaeva, Ekaterina / Baiburova, Olga / Manzhula, Oksana (2022): „Form – Meaning – Usage Synergy in LSP & Professional Communication: Computer Security in Terms of Frame Semantics." *Fachsprache. Journal of Professional and Scientific Communication* 44.3–4: 169–191.

We have compiled and examined a collection of the CS terms and a corpus of LSP texts contextualising the terms. The database is built as an integral corpus of texts centred around selected terms. This is relevant to take them inseparably, for the terms' semantic frames can be reconstructed only within their context. For this reason, in this paper, we will refer to this combination as the corpus of terms. We collected the corpus within a project on an interdisciplinary terminological dictionary development. The project is ongoing at the Department of English for Professional Communication of Perm State University and is included in an ESL course for non-linguistic faculties. The second-year CS students have collected the corpus used for our research. During the course (1 academic year), the students had a monthly task to read contemporary specialised texts, including journal articles, books, and documentation on computer virology and select words, which had a specific meaning in CS. The students worked in one document shared for editing via Google sheets to collect unique terms. The terms have been stored with their contexts and supplied with specialised definitions by CS experts. The terms have been sorted into predefined thematic categories according to their contextual meaning, i. e., the meaning realised in the context of a professional communication event: *Virus type, Malicious activity / Malefactor, Software, Hardware, Vulnerability, Operating systems, Safeguard measures, Computer networks, Mathematics / Functions, Data,* and *Programming languages*. We have selected only nominative terms, i. e., nouns and noun groups. A noun group is "a group containing at least one noun or pronoun (the head) and often other items such as determiners, adjectives, and prepositional phrases" (The Free Online English Dictionary), e. g., *computer virus*. So, the nominative terms made up a subcorpus of 355 units.

We aim to analyse which semantic roles (SRs) are prevalent for CS terms in context. We use the semantic frame modelling method, which consists in assigning Fillmore's SRs to the participants of an event or situation in professional communication.

2 Background knowledge and current vistas of semantic framing

Katsnelson believed that the word form and the mental content are given in the language in a complex and contradictory unity. To reveal their dialectics means to trace the transitions from the meanings of words to concepts and from grammatical categories to categories of thought (Katsnelson 2010: 397). This idea correlates with the theory of a bilateral sign, i. e. the relationship between 'the signifier' (a linguistic form) and 'the signified' (the meaning of the form) (Saussure 1959). The dichotomy of language and thinking was examined by von Humboldt, who believed that the forces that generate language and thought are inseparable (Humboldt 1984: 305), and Sapir-Whorf, who stated that a person's picture of the world is primarily determined by the system of the language he speaks (Whorf 1956).

The language grammatical system preserves the structural nature of human consciousness and reflects the natural world in a folded and syncretic state (Solomonick 2011: ii–iv) formalised to semantic frames. The frame comprises elements assigned with standardised SRs (Fillmore 1971). We apply this logic to specialised texts, which preserve the grammatical structure of the language but are filled with specialised lexics, capable of evoking abstract descriptions of professional situations through semantic frames (L'Homme 2017: 8).

Fillmore's theory has been comprehensively studied and implemented in computational linguistics. The theory's potential is evident in natural language processing and machine learning due to its aptitude for standardisation and categorisation. Furthermore, the labelling of SRs provides an easy way to conceptual modelling and, thus, conceive the logic and the mechanism

of thinking. Due to its rule-based nature, this ability is scaled to artificial thinking, crucial for natural language processing and machine learning.

The task of mapping word tokens to frames they evoke, and for each frame, finding and labelling its argument phrases with frame-specific SRs, is well developed on the technical side. However, there is still a problem related to a small amount of manually pre-trained data for supervised training to achieve precise automatic parsing (Kshirsagar et al. 2015: 218).

Impressive results in frame semantics have been achieved by Faber/Cabezas-García (2019) and the LexiCon Research Group (2021), who elaborated efficient methods for parsing, lexicon building, and semiautomated extraction of metaphor-related terms in the environmental domain. However, the domain of CS lacks well-elaborated automatic semantic parsing, and there is a demand for manual semantic research underpinning automatic parsing.

3 Proving evidence for form – meaning – usage synergy

3.1 Logic of the analysis

SRs are the roles that a noun phrase (NP) may play with respect to the action or state described by a governing verb (V), commonly the sentence's main verb. An SR is "a part of the predicate semantics that reflects the general properties of the predicate argument" (Plungian 2003: 3). For instance, in 'a virus infected the computer', *infected* is the predicate, while *virus* and *computer* are the predicate arguments. Establishing relations between the predicate and its arguments allows for formal semantic analysis of the meaning underlying the utterance based on the SRs labelling of NP+V units. To illustrate how meaning unveils in the discourse, i. e. in usage, and is determined by grammar and syntax represented in SRs, i. e. form, we apply semantic framing on our sample corpus. The labelling has been done manually to all the term parts of NP+V units regarding the basic features of the SRs.

We concentrated on the SRs traditionally singled out in the frame semantics: Agent, Counteragent, Objective, Perceptive, Cause, Benefactive, Addressee, Patient, Result, Locative, Trajectory, Instrument, and Goal. Additionally, we suggest new SRs, namely Specifier (cf. section 3.2) and quasi-Agent (cf. section 3.14), especially relevant for specialised discourse. Since we are interested in the pragmatic conditionality of word meanings, we apply the principle of SR's labelling to the terms in their contexts and infer the interdependence between the SRs and the terms' categories. So, the logic of our analysis includes the following steps: 1) an SR description; 2) sample analysis, i. e., determining an SR in the term's NP+V unit based on their semantic meanings retrieved from dictionary definitions; 3) inferences on the role–category interrelations. To obtain the meaning of the term, we used official corpus-based dictionaries, such as "The Free Online English Dictionary" from Macmillan Publishers and "The Longman Dictionary of Contemporary English Online"; the collection of dictionaries by subject, e. g., "The Free Dictionary" by Farlex or other published professional sources (books, articles, etc.) and professional IT and CS encyclopaedias or fora, e. g. "Techopedia" defining novel terms not registered in the official dictionaries.

3.2 Pragmatic potential of the Agent

The Agent is one of the leading SRs, an active participant, i. e. a person, a subject, an animated pathogen, or a natural force performing an action or exercising control over the situation

(Fillmore 1968: 24, Cook 1998: 5). Such nouns as *heat* and *wind*, being inanimate but representing objects that can act, are also considered the Agents (Chafe 1970: 7). In CS, the Agent is generally introduced by inanimate nouns, which designate entities capable of operating on themselves or others, "usually to bring about some change in the location or properties of itself or others" (Downing/Locke 1992: 5). Thus, the Agent's typical characteristic features, such as animation, intent, motivation, and responsibility, are attributed to inanimate professional concepts.

In our corpus, the Agent comprises the terms belonging to the *Malicious activity / Malefactor* category (11 cases). The less represented categories include *Virus type, Software, Hardware, Vulnerability, Operating systems* (Figure 1).

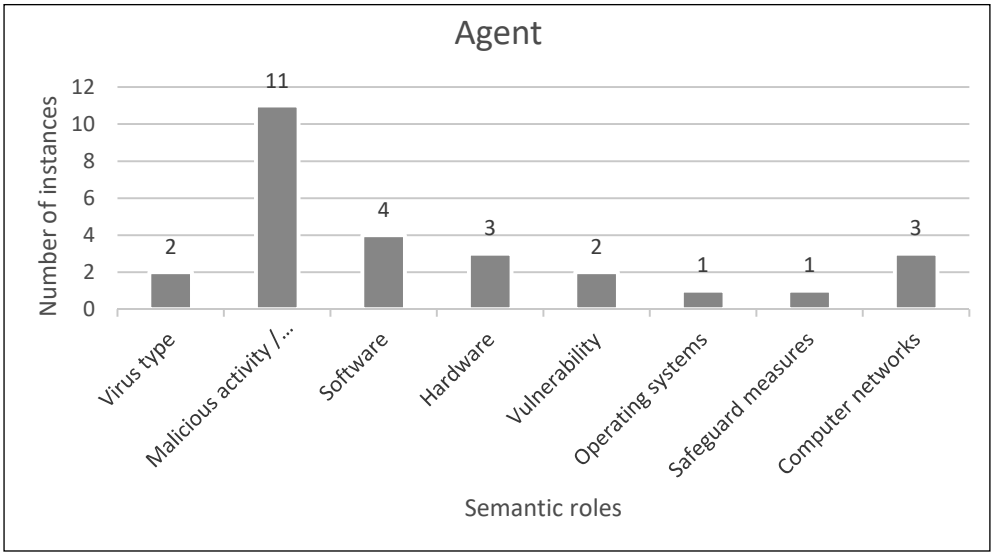

Figure 1: Distribution of the Agent role in the CS discourse

An example of the Agent played by the term designating a virus type is given in (1):

(1) The *Trojan* keeps records of the checksums for the obtained data. (Mamedov/ Sinitsyn 2016)

Sample (1) illustrates the capability of *the Trojan*, i. e. "a program that seems useful but is designed to be harmful, for example by stealing or destroying information" (The Free Online English Dictionary), to perform intelligent actions such as *keeping records*.

The verbs expressing the Agents' actions contain the semes of animacy, activeness, and deliberateness:

- *attempt*, i. e. "to make an effort to achieve or complete (something difficult)" (English Dictionary, Thesaurus & Grammar Help),
- *download*, i. e. "to move information to your computer from another computer system or the Internet" (The Free Online English Dictionary),
- *bypass*, i. e. "avoid dealing with someone or something, especially because you think you can achieve something more quickly without using them" (The Free Online English Dictionary),

- *target*, i. e. "intend or try to attack someone or something" (The Free Online English Dictionary),
- *keep records*, i. e. "to regularly record written information somewhere" (The Longman Dictionary).

All these activities are typical of humans. They comprise the idea of reaping the benefit, putting it into practice, implementing, etc. These verbs designate the activity typical of malware. They have negative connotations and deliver the meaning of over-persuading, instigating, and involving in some troublesome business.

Terminological categories of *Malicious activity / Malefactor* and *Software* designate the programs, whose prototypes, i. e. their biological counterparts, are pathogen agents, propagating, spreading, and causing infectious diseases. All these imply agency as the ability to act as the initiator of some action.

Although SRs are primarily concerned with the semantic relations of the arguments and their predicates, their syntactic relations cannot be neglected. These relations are described through the argument's position relative to its predicate and the part of the sentence the argument occurs in. This is particularly relevant if the study results are to be used in automatic parsing. Suppose semantic and syntactic properties of the instigator of the action disagree, e. g. in sentences with the passive structure. In that case, it makes sense to think of introducing quasi-roles[1], here quasi-Agent, which will contribute to the precision of the data description and rules formulating to achieve a higher quality of automatic parsing. In our corpus, the terminological categories that represent this quasi-role include *Malicious activity / Malefactor* and *Software*, for example:

(2) Spain and Poland have been two countries traditionally targeted by *SMS scams* and similar *malware*. (GReAT 2017a)

The term *SMS scam* refers to malware ("any software that brings harm to a computer system" [Techopedia]), which occurs when cybercriminals use false text messages asking customers to provide personal or financial information. Both terms (*SMS scam* and *malware*) execute the action denoted with the verb *target*, typical of the Agent. The preposition *by* shows that the participant following it is active and is the initiator of the action. Yet, the originally inanimate nature of the term and its syntactic role in the passive structure make us doubt a pure Agent role of *SMS scam* and consider it from the prism of quasi-realisation. To define the quasi-Agent role, we highlight its right-handed position regarding the predicate, the presence of the preposition *by*, preceding the argument designating the doer of the action, expressed by the predicate.

3.3 Pragmatic realisation of the 'shadow Agent'

The next SR of the Counteragent is a participant in a situation that qualifies through a counteraction relationship. The "shadow Agent" (Paducheva 2004: 361) implies "the force or resistance against which the action is carried out" (Fillmore 1971: 376) or "a substance that impedes the commission of an action" (Gak 1998: 413).

We define the Counteragent as a right-handed argument of the predicate, expressing a

[1] Here, the idea of quasi-roles is given as possible solution to the problem of syntactic-semantic mismatch in the roles distribution within the professional discourse. The aspect needs further research.

computer or cyberworld entity, action, or process which enters into counteraction relations with the Agent, i. e. the Agent acts against or prevents something unwanted expressed through the Counteragent.

In the Counteragent role, the terms from the categories of *Malicious activity / Malefactor, Safeguard measures, Software,* and *Operating systems* can be found (Figure 2).

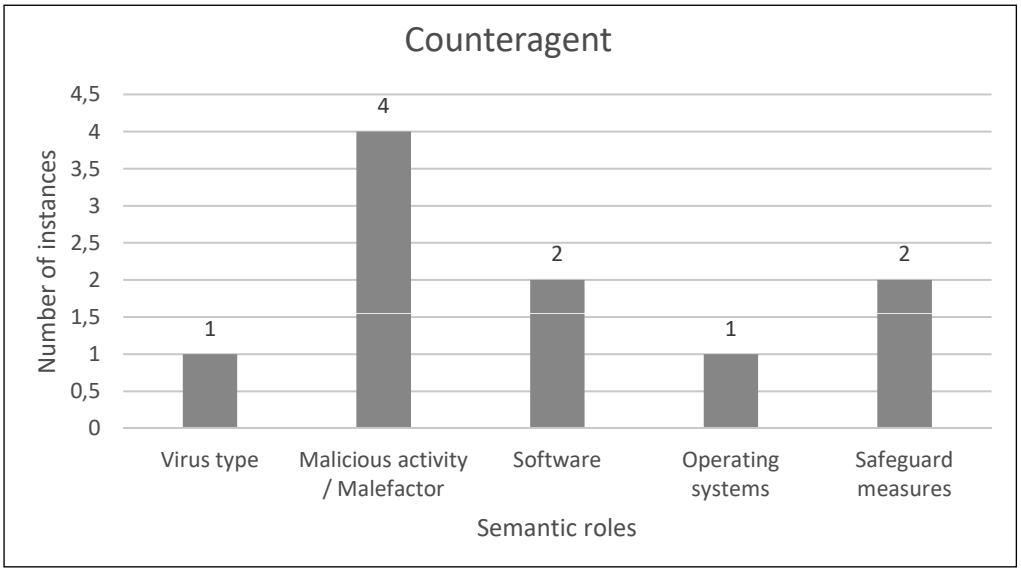

Figure 2: Distribution of the Counteragent role in the CS discourse

The most frequent category is *Malicious activity / Malefactor* exemplified in (3):

(3) If the victim successfully combats *SYN-flood*, the attacker can switch the scenario on the control panel and evaluate the victim's reaction. (Makrushin 2017)

The term *SYN-flood* stands for "a type of network or server degradation attack in which a system sends continuous SYN requests to the target server to make it overconsumed and unresponsive" (Techopedia). This action is malicious, for it causes soft- or hardware malfunction or information security breach. When taking countermeasures against such kinds of activities, the most likely verbs to deliver them are

- *combat*, i. e. "do something in order to try to stop something bad from happening or a bad situation from becoming worse" (The Free Online English Dictionary),
- *prohibit*, i. e. "officially stop something from being done, especially by making it illegal" (The Free Online English Dictionary),
- *disable*, i. e. "deliberately make a machine or piece of equipment impossible to use" (The Longman Dictionary),
- *attack*, i. e. "deliberately use violence to hurt a person or damage a place" (The Longman Dictionary).

These verbs imply carrying out deliberate actions against something, being part of countermeasures. They also comprise the meaning of preventing something from happening, fighting or destructing. The actions are usually executed against something unwanted.

3.4 Pragmatic diversity of the Objective

The Objective is an item, "the action is directed to" (Cook 1998: 5), affected by the action (Fillmore 1968: 25). The Objective usually acts as a direct complement and "the object that is exposed by the verb" (Downing/Locke 1992: 5). All these features bring us to the following definition of the Objective – a right-handed argument of the predicate expressing a computer or cyberworld entity manipulated by someone or something defined by the argument. However, the entity does not change or cease existing as a result of this manipulation.

In CS, the Objective is rather heterogeneous regarding terminological categories (Figure 3):

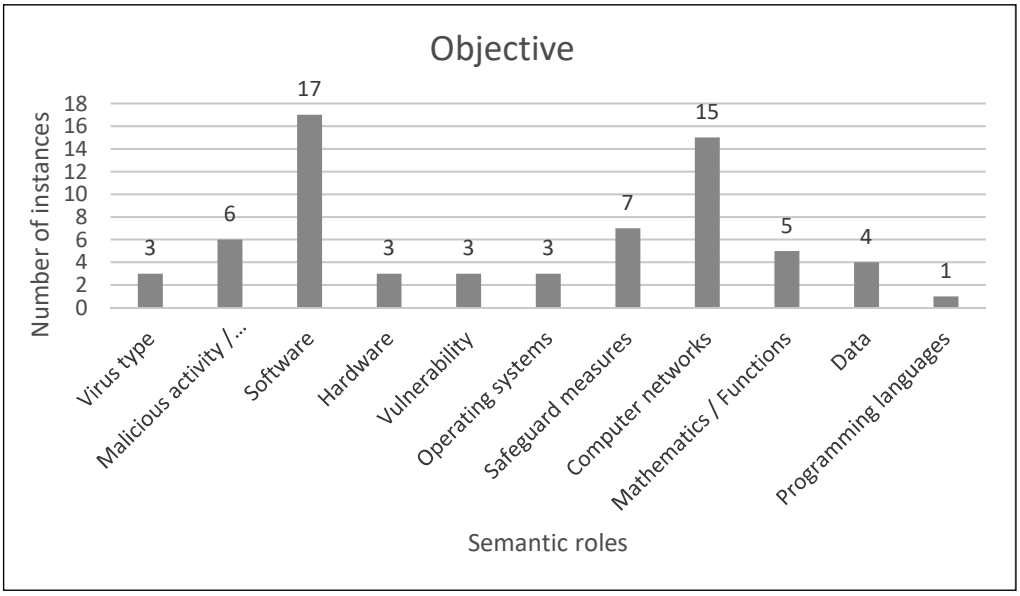

Figure 3: Distribution of the Objective role in the CS discourse

The most typical categories are Software and Computer networks. The Objective is often implemented by terms denoting: mechanisms, tools, operations, network clients, software, and malware, like *malicious code*, i. e. "a code causing damage to a computer or system" (Techopedia), as in (4):

(4) A vulnerability is a fault in a program's implementation that can be used by attackers to gain unauthorised access to data, inject *malicious code* or put a system out of operation. (Zakorzhevsky 2015)

Among the verbs conveying this SR are *include*, *receive*, *download*, and *check*. They are used with the terms of the *Software* category due to the common seme of object manipulation inherent in these verbs. They imply moving something from one location to another, holding something, and the presence or quality testing.

The categories of *Software* and *Computer networks*, primarily referring to the software part of the networks, are significant for the Objective because programs are engaged in different kinds of activities, including the network ones, carried out by some other participants but do not usually undergo any changes.

3.5 Tangibility of the computer security Perceptive

The Perceptive is "an integral semantic attribute for right-handed arguments of most sensory verbs" (Amirova 2002: 119). This SR indicates that the object is perceived by the Agent through physical senses. This act of perception usually evokes emotions or cognitive change in the Agent.

The Perceptive is provided by the terms of *Virus type, Malicious activity / Malefactor, Software, Hardware, Vulnerability, Operating systems, Safeguard measures,* and *Computer networks* (Figure 4).

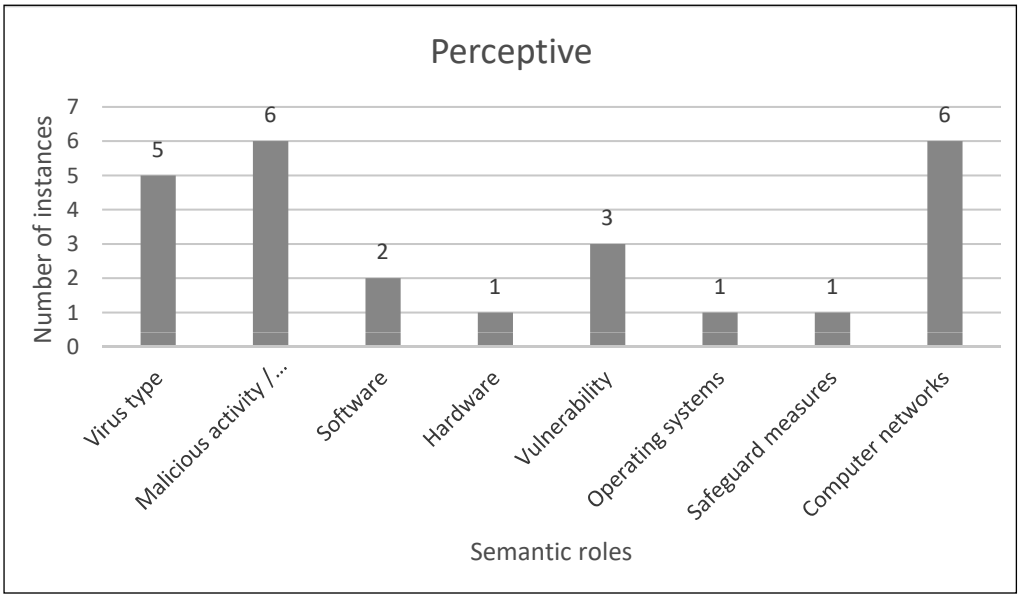

Figure 4: Distribution of the Perceptive role in the CS discourse

The most substantial categories are *Malicious activity / Malefactor, Computer networks,* and *Virus type*. The Perceptive is played by the following subcategories: kinds of law breach, techniques or attack vectors, computer programs, attack scenarios, and processes, like *leak*, i. e. "the origin of secret information that becomes known, or the act of making it known" (Cambridge Dictionaries Online), as in (5):

(5) These are focused on analysing single apps to detect information *leaks* through inter-component communications, ICC. (Blasco/Chen 2018)

The verbs that appear in CS texts to convey the relations between the Agent and the Perceptive are *detect*, *identify, consider, hear about*, etc., which mean perceiving something, making it evident, and finding. The Perceptive can also be highlighted by the words *remind*, *like*, *similar to*, *under the veil*, and others, which express comparison with some object or process, determining identity, matching unique features. These words signal the attempt of the entity behind the Agent to precept or cognise the entity expressed by the Perceptive.

The dominance of the categories of *Malicious activity / Malefactor, Computer networks,* and *Virus type* can be explained by the fact that the CS concepts designated by the terms of these categories are still to be understood, recognised, and conceptualised by the user, who can

do nothing but observe the malicious acts, which are primarily executed in or via computer networks.

3.6 Connotative specificity of the Cause

The Cause's function is "to lead to a change in the state of Affected Participant" (Downing/Locke 1992: 25). The participant in the Cause role expresses the reason for the state occurrence or change.

Though the Cause is exemplified by the terms belonging both to the categories of *Malicious activity / Malefactor* or *Virus type* and *Safeguard measures* or *Software*, it always bears a negative connotation. In (6), *OPSEC* (*Operations Security)* involves the identification and protection of generally unclassified critical information or processes that a competitor or adversary can use to gain real information when pieced together (Unuchek 2017). It designates the process thought to safeguard data. However, in practice, it contributes to the growth of illicit business and crime:

(6) Due to its robust anonymity, *OPSEC* techniques, low prices, and client-oriented strategy, the Dark Web remains an attractive medium for conducting illicit businesses and activities. (Unuchek 2017)

Consistency checks in (11) denote "a test performed to determine if the data has any internal conflicts" (Computer Hope) but, due to their co-occurrence with the negatively connotated adjective *insufficient,* represent the cause of unwanted actions:

(7) He has reported a number of serious vulnerabilities: Remote Code Execution from web scripts, arbitrary device firmware modification due to insufficient *consistency checks* ... (Threat intelligence report for the telecommunications industry 2016)

The Cause markers are the prepositions *due to* and *because of*, which indicate a cause-and-effect relationship between the participants and the events and express neutral and negative meaning. This property correlates with another feature of the Cause mentioned above, i. e. the usage of adjectives like *insufficient*, *poor*, *inadequate*, *wrong*, *improper,* or the terms, like *Trojan*, *malware*, *adware*, both with a negative connotation. The Cause helps the reader understand why a particularly adverse effect occurs.

3.7 The search for the Benefactive of the computer security issues

The Benefactive, means "the possession of an object with state verbs or a participant in the transfer of information with procedural and action verbs" (Cook 1998: 151). The Benefactive is "an object for which an action is performed" (Downing/Locke 1992: 152), "a person or object that receives something as a result of an action" (Brinton 2000: 82), but "it is not necessary to receive benefits" (Downing/Locke 1992: 152). Beneficial verbs denote the possession or transfer of property. Despite their generally positive connotation, the case, often associated with the Dative, can be positive and negative, i. e. a person can gain or lose the property. This fact brings us to further deliberations, outside the scope of this paper, about splitting this SR into three, namely expressing positive, negative, and neutral influence. In our corpus, the Benefactive occurs with such verbs as *have, possess, inherit, give, fetch, buy, cook (make),* etc.

The Benefactive is carried out by the terms of *Malicious activity / Malefactor, Software, Safeguard measures,* and *Computer networks*. The distinctive feature of the SR is the preposition *for*, which helps to express the idea that some actions open new functions to the notions following this preposition. In CS, the implications might be either positive or negative for the user. In (8), the work of a particular method enables *Internet Explorer* to become a medium for malicious activity:

(8) Since this method only works for *Internet Explorer*, the malware needs to force the user to access internet banking via that browser. (Marques 2016)

The fact that the Benefactive is exemplified by the terms of the categories *Malicious activity / Malefactor, Software,* and *Safeguard measures,* and *Computer networks* can be explained by the peculiarities of the CS discourse, where a lot of efforts are taken to improve computer and communication protection or, vice versa, malware.

3.8 Pragmatic function of the Addressee

The Addressee is the person to whom the action is directed (Apresjan 1995: 25). In the Addressee role, we have identified only two terms belonging to the categories of *Computer networks* (14) and *Malicious activity / Malefactor* (15). This SR is played by the terms, designating network software, like *remote server* in (9) or *the money mule's cell phone number* in (10).

(9) The stolen Paypal credentials were forwarded to another *remote server* located in Mexico. (Naor/Alon 2016)

A remote server is "a server that is dedicated to handle users that are not on the LAN but need remote access" (What is a remote server 2017). This term belongs to the Computer networks category. In (9), *the remote server* acts as a personified addressee to which an object (here *credentials*) is forwarded. Personification and anthropomorphism of software and hardware is a widespread phenomenon found throughout in the CS discourse (Isaeva/Baiburova/Manzhula 2022). The *remote server* occurs in the frame, which evokes a mental construction of the event of human-to-human interaction, namely forwarding something from one person to another. This is a case of the metaphor representing a behavioural comparison.

(10) The attacker issues a money transfer to *the money mule's cell phone number*. (GReAT/Naor 2016)

Here, we witness an example of anthropomorphism in the CS concepts, which arises within the 'transfer to' frame. The Addressee is the money mule's cell phone number, i. e. the cell phone number of a person allowing their account to be used to receive fraudulent funds and then withdrawing the money on behalf of a fraudster ("Gang of fraudsters and 'money mules' sentenced for £200k scam" 2019). The CS event triggers the metonymic transfer of the money mule's intimateness and their SR to the number of cell phones used in the malicious activity. In the frame of a typical CS event, the Addressee is attributed to some device, hardware, a user, or any victim of a malefactor which receives some malware or fraudulent item.

3.9 A computer security Patient as the Object exposed to change

The Patient is "the recipient the impact is directed to and whose physical state, including position in space, changes as a result of this situation" (Fillmore 1968: 68). It refers to "a person or object exposed and undergoing a change" (Brinton 2000: 22). This SR is illustrated by the groups of terms determining *Software, Operating systems,* and *Computer networks* since they are the primary targets for manipulation and change by the malefactor using the malware, e. g. (11):

(11) The attackers try to avoid an early detection due to wrong timeserver settings, since the current *NTP Server entry* will be overwritten by the previous malicious requests. (Ortloff 2016)

NTP Server entry is a record of the server location, which refers to a "protocol used to synchronise computer clocks across data networks" (Techopedia). The term denotes data communication rules exposed to manipulation and change. They can be expressed in language by such verbs as

- *configure,* i. e. "to arrange something or change the controls on a computer or other device so that it can be used in a particular way" (Cambridge Dictionaries Online),
- *overwrite,* i. e. "replace a computer file with a different one" (Cambridge Dictionaries Online),
- *disable*, i. e. "permanently or temporarily turn off" (Computer Hope).

The verbs' semantics contains the semes of change and influence. The terms of the groups of *Software, Operating systems,* and *Computer networks* occur as the Patient for they determine computer items prone to manipulations, unable to influence the event causing their alterations.

3.10 Types of the Result of the computer security activities

The Result is "an object or creature arising from an action" (Fillmore 1968: 25). The SR comprises the terms of *Virus type, Malicious activity / Malefactor, Software, Operating systems, Safeguard measures, Computer networks, Data* (Figure 5).

The distinctive groups are *Software, Virus type,* and *Malicious activity / Malefactor*. The terms can designate attacks, forms of cybercrime, computer operations, and software, e. g. (12):

(12) They created an illegal *add-on* to the legal RBS product. (Stoyanov 2016)

The Result is delivered with the verbs:

- *create*, i. e. "to make something new or original that did not exist before" (The Free Online English Dictionary),
- *implement*, i. e. "to carry out; put into action" (Collins),
- *develop*, i. e. "to invent something or bring something into existence" (Cambridge Dictionaries Online),
- *cause*, i. e. "to produce a result" (Your Dictionary),
- *perform*, i. e. "to do an action or piece of work" (Cambridge Dictionaries Online).

These verbs contain the seme of producing something.

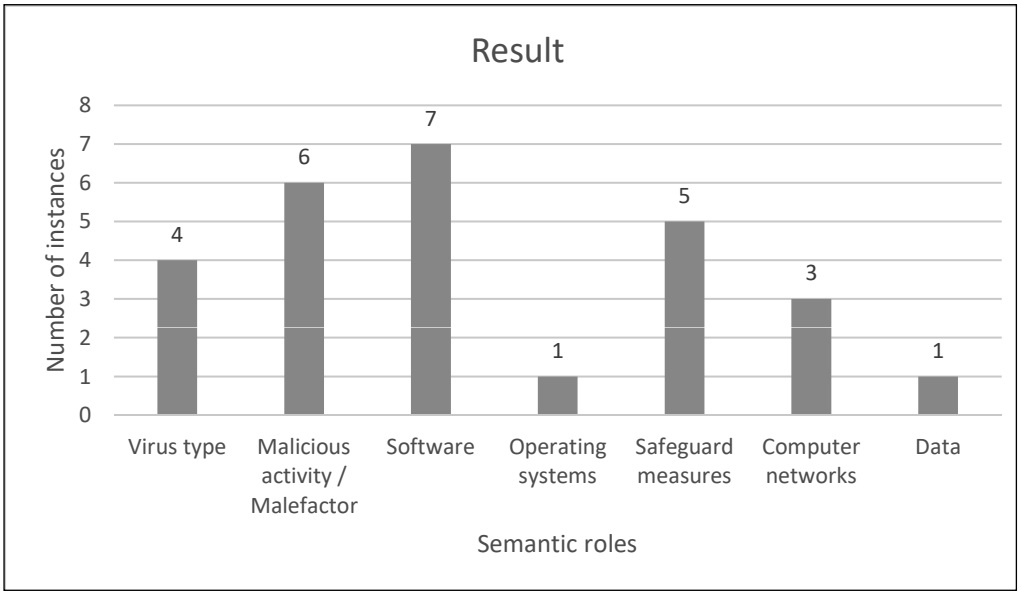

Figure 5: Distribution of the Result role in the CS discourse

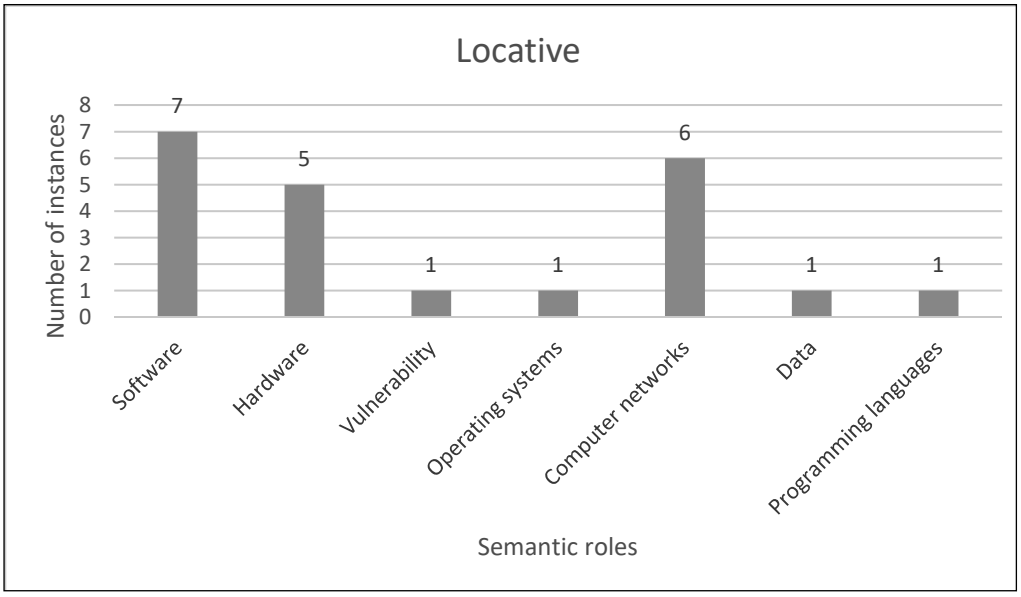

Figure 6: Distribution of the Locative role in the CS discourse

3.11 Spatial orientation: Locative

The Locative is the place of action, physical location, or spatial orientation (Fillmore 1968: 25; Cook 1998: 127). This case has two types, indicating the state and direction.

The Locative is provided by the terms of *Software, Hardware, Vulnerability, Operating systems, Computer networks, Data,* and *Programming languages* (Figure 6).

Most of the terms determine computer networks, protocols, traffic, types of websites, terminals, and parts of a computer; e. g. *hard drive* in (13), i. e. "the part of a computer where information and programs are stored, consisting of hard disks and the electronic equipment that reads what is stored on them" (The Longman Dictionary):

(13) The malicious program was unusual. Unlike most other malware, it left no traces on the *hard drive* of the system attacked and worked only in the RAM of the machine. (Stoyanov 2016)

The Locative verbs include *appear, leave traces*, *write*, *find*, and *work*. Their semantics contains the realisation of some activity in a particular place, namely some electronic or digital environment.

3.12 Spatial orientation: Trajectory

The Trajectory denotes "the path in which they move from one place to another in the process of action" (Brinton 2000: 68). As the Trajectory, the terms of *Hardware, Operating systems,* and *Computer networks* appear (Figure 7).

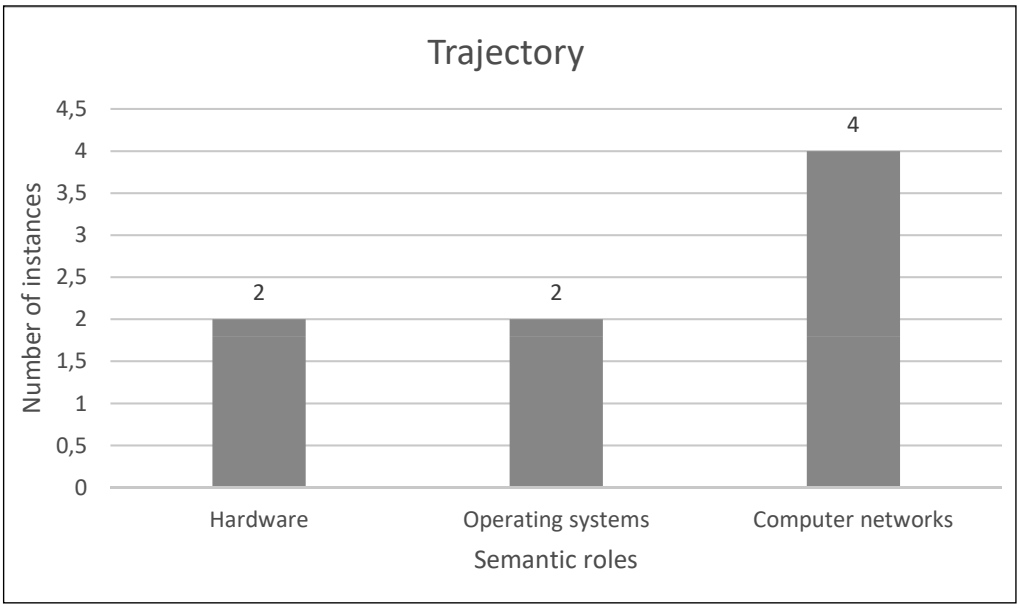

Figure 7: Distribution of the Trajectory role in the CS discourse

In (14), *inter-component communications (ICC)* determine the mechanisms forming "the basis of a broader environment designed to support the construction of educational applications

[…] constructed by end-users […] by assembling high-level, domain-specific software components into functional wholes" (Koutlis et al. 1998):

(14) These are focused on analysing single apps to detect information leaks through *inter-component communications, ICC.* (Blasco/Chen 2018)

The Trajectory is delivered with the prepositions *through*, *via*, *along,* and *over,* introducing the medium or channel for program or data flow transmission.

3.13 Spatial orientation: Goal

The Goal denotes where some software or malefactor gets access to, or some file or data are uploaded/downloaded to/on. This SR is low-frequent in our dataset. We have identified ten cases in *Safeguard measures, Computer networks, Malicious activity / Malefactor, Software, Hardware,* and *Data* (Figure 8).

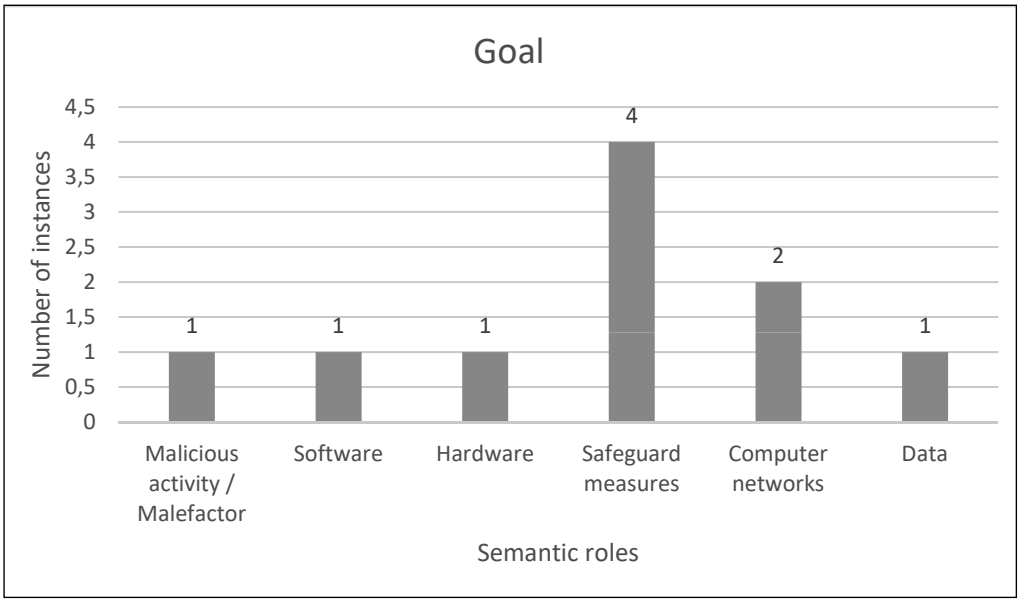

Figure 8: Distribution of the Goal role in the CS discourse

For example, the *Safeguard measures* category contains terms for types of data encryption, passwords, and other security technologies:

(15) After infecting their victims with banking malware and obtaining their phone numbers, they called the CSP's support and […] asked for a new SIM card to be activated, thus gaining access to *OTP.* (Threat intelligence report for the telecommunications industry 2016)

In (15), *OTP (One-Time Password)* stands for "an automatically generated numeric or alphanumeric string of characters that authenticates the user for a single transaction or session" (Techtaget 2021). This numeric string became a goal for banking malware whose activity is expressed with *gain access to*.

3.14 Descriptive function of the Specifier

To elaborate the descriptive function of the forward terms in noun phrases, i. e. noun+noun phrases, we introduce the Specifier. The Specifier complements another argument and consequently is not directly related to the predicate but adds new qualities to the main argument in the noun phrase. This SR is typical of the English language for the nouns in this SR carry out the attributive grammatical function and help to determine particular types of safeguard measures, commands, instructions, computer networks, software, or malicious activity, e. g. *spoofing,* which stands for "hacking or deception that imitates another person, software program, hardware device, or computer, with the intentions of bypassing security measures" (Computer Hope):

(16) NBNS is vulnerable to *spoofing* attacks. (Assolini/Makhnutin 2013)

Spoofing determines the type of NBNS vulnerability.

A Specifier can occur after a specified noun. In this case, the Specifier is introduced with the preposition *of*, as seen in (17):

(17) This threat was originally discovered by a bank's security team, after detecting Meterpreter code inside the physical memory of *a domain controller*. (GReAT 2017b)

Here *a domain controller*, i. e. "a server that responds to security authentication requests within a Windows Server domain" (Techopedia), specifies compromised hardware, particularly the physical memory which hosts malicious software.

This SR might occur in any thematic category (Figure 9):

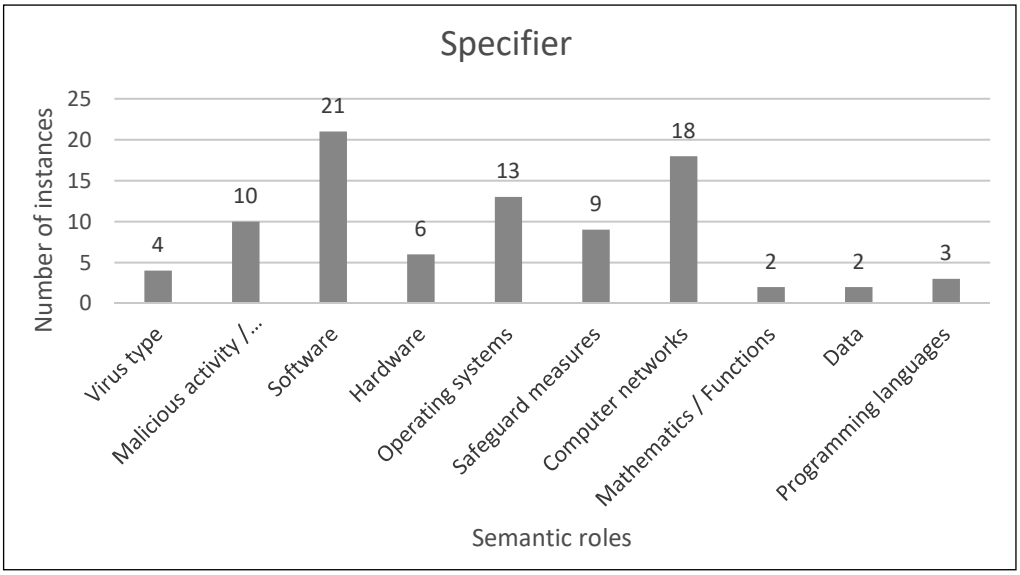

Figure 9: Distribution of the Specifier role in the CS discourse

The Specifier at the cognitive level helps establish the structural organisation of the device and assume possible adverse effects.

3.15 An instrumental part of the computer security activities

The Instrument is an "inanimate force or object involved in an action" (Fillmore 1968: 25), "a mean by which an event is raised, or a tool, usually, an inanimate one used to carry out an action" (Brinton 2000: 168). The Instrument is realised by various term groups – *Malicious activity / Malefactor, Software, Hardware, Operating systems, Safeguard measures, Computer networks, Mathematics / Function, Data,* and *Programming languages* (Figure 10).

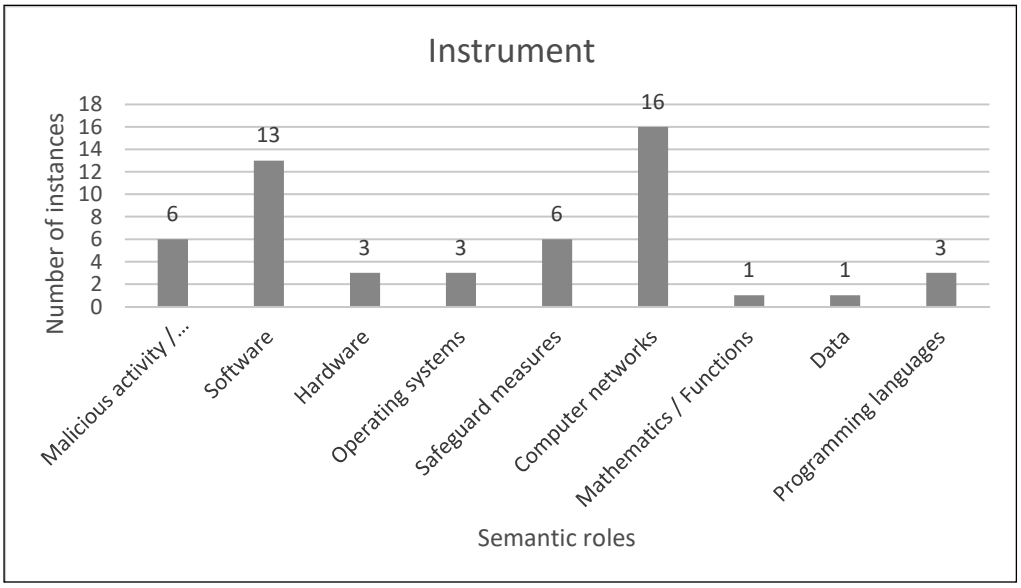

Figure 10: Distribution of the Instrument role in the CS discourse

The most frequent groups are *Computer networks* and *Software*. The terms often designate concepts that can be subdivided into three main groups:

a) Using software for data protection, e. g. *AES*, i. e. *Advanced Encryption Standard,* which is a symmetric-key block cypher algorithm and U.S. government standard for secure and classified data encryption and decryption (Techopedia), e. g. (18):

(18) The files are encrypted using *AES* with CBC mode. (Naor/Alon 2016)

In (18), the Instrument mode helps the reader understand what tool has been used to secure data.

b) Using malware, e. g. *Lurk* in (19) for data violation:

(19) At that time, the "company" had two key "products": the malicious program, *Lurk*, and a huge botnet of computers infected with it. (Stoyanov 2016)

The noun, executing the Instrument, *Lurk,* designates a versatile malicious computer program of a Trojan type, which "can steal money from bank customers" (Shulmin/Prokhorenko 2016). The preposition *with* introduces the Instrument of the pronoun *it* (*Lurk*). This SR helps infer

the latent Agent for better understanding that malware is only a tool used by a plotter to execute malicious actions.

c) Using software as a vulnerability for a data violation. Example (20) represents a CS conundrum when software initially developed for CS protection is used as a tool for malicious actions:

(20) After using *anti-rootkits* Brazil's cybercriminals went deeper and started to develop their own bootloaders. (Marques 2016)

Here (in 20), *anti-rootkit*, i. e. "a tool designed to identify various threats like a rogue and suspicious processes, hooks or modules, registry keys, modified files, and known / unknown rootkits" (Lad 2011), becomes a vulnerability that can be abused. The Instrument is introduced with the verbs *use* and *utilise* and the preposition *with*.

4 Results

We have analysed the SR distribution in the context of CS terms in professional communication. The terms have been sorted into 11 categories. The categories have been devised collaboratively by cognitive linguists and CS experts. Employing domain experts in the project is an effective way to overcome the problem of low lexicographers' and terminologists' expertise in the field (L'Homme 2017: 11). At the stage preceding data collection, the choice of the categories was motivated by the experts with reference to the aspects covered in the CS courses. Subsequent semantic analysis of each term in the corpus and the continuous database enrichment with new terms caused further finetuning in the categories. Thus, the categories are domain-specific. They generally incorporate information on the main participants of the CS events, types of their interaction, typical settings and media or environment, methods and techniques used in the domain, etc. The terms from the selected corpus have been assigned SRs based on their contextual meaning. The statistic shows the SRs prevailing in our dataset (Figure 11).

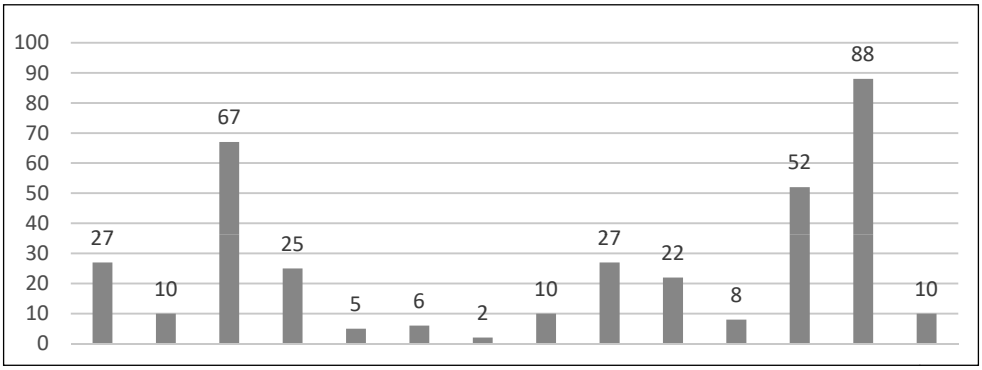

Figure 11: SR metrics in the dataset

Figure 11 shows that the most numerous SRs are Specifier, Objective, and Instrument. The Specifier can hardly count as typical of only the CS discourse because it is common for English grammar in general. High frequency of the Objective was also expected for a specialised discourse, where manipulation with objects is a routing business. However, in CS, the objects are

primarily digital, thus, intangible. Specific to our dataset is the actualisation of the SRs with a frequency rate of 52 and below.

We have determined a certain dependence of SRs on the terms' thematic groups.

- Thus, the **Agent** is enacted by the terms denoting participants able to initiate or cause some action, as a rule, a malicious or defensive one. The difference from the non-specialised discourse is that the terms acquire animacy and exhibit metaphoricity. This is due to the essential property of the Agent to label the animate instigator of the action.
- In professional communication, passive constructions, in which the initiator of action is presented indirectly, are frequent. Such a participant does not occupy the subject position in the sentence and is introduced with the preposition *by*. That is why a new SR of the **quasi-Agent** has been added to designate an activity implementer in passive constructions and back syntactic and semantic synergy.
- The **Counteragent** is the Agent's antipode. It represents the force or the initiator of the force, against which the Agent's effort is directed. Accordingly, in our corpus of CS texts, the most numerous group for this SR has been *Malicious activity / Malefactor*.
- The **Objective** represents the most prominent feature of inanimateness attributed to computer entities brought to action by computer experts or users. The Objective occurs in almost any category and designates computer tools, programs, network clients, mechanisms, etc. However, the most illustrative category is *Software*. The category includes computer programs engaged in different kinds of activities carried out by other participants who operate these programs but do not change them.
- The **Perceptive** is also typical of the categories determining entities that do not perform any self-activity but occur as objects seen, heard, or sensed by other participants. The SR is regular of the *Malicious activity / Malefactor* terms. In many cases, users become aware, passively observe or express some attitude to malicious actions in the computer sphere but cannot influence the situation.
- As the **Cause**, there occur the terms for programs or activities generating a problem for CS. They mainly refer to *Malicious activity / Malefactor, Virus type, Software,* and *Safeguard measures*. The latter proves that cybercriminals take advantage even of CS measures.
- The **Benefactive** is typical of *Malicious activity / Malefactor, Software,* and *Safeguard measures*. The terms in this SR designate computer entities, which get new features or qualities from other computer entities.
- The **Addressee** is mainly found in the categories of *Computer networks* and *Malicious activity / Malefactor*, which contain terms for network software, device, hardware, or a computer user, who receives malware.
- The high frequency of the **Patient** in such categories as *Software, Computer networks, Data*, and *Operating systems* proves that the entities designated by these terms often suffer from malware and are modified because of malicious actions.
- The **Result** is attributed to what usually emerges from the fraudulent situations, the most relevant category being *Malicious activity / Malefactor*.
- The place where malware is likely to be found is determined by the **Locative** or the **Trajectory**. They are nominated by the terms from *Hardware, Computer networks, Software,* and *Operating systems*.
- And finally, to identify the SR of terms that define the features of a CS item, we have

added the **Specifier**. It is situated immediately before the specified word or after it. In the latter case, it is preceded by the preposition *of*.

So, all the basic SR have been identified in the CS discourse, and their attribution to particular thematic categories has been reasoned.

5 Discussion

The results can be interpreted through the philosophy of sign systems. LSP representing a separate branch of natural language, follows the logic attributed to any sign system. According to Solomonick (2011), this logic occurs in three types. The first one is the logic of the correspondence between the language system and reality. This means that the language reproduces events to preserve the same relations and dependences as in real life. To illustrate this assumption, one should refer to the semantic frame, which mirrors scenarios of everyday real-life situations. Thus, the SRs and their valences match real-life stereotypical participants entering similar relations (21):

(21) *Colluding apps* bypass the security measures enforced by sandboxed operating systems such as Android. (Blasco/Chen 2018)

Colluding apps play the Agent SR. They change the trajectory of their movement not to come in contact with (i. e. *bypass*) an unwanted obstacle (i. e. *security measures*, which play the Objective SR). *Sandboxed operating systems such as Android* play a double SR – the Counteragent with respect to *colluding apps* and the quasi-Agent with respect to *security measures.*

To understand this situation typical of the CS discourse, if one does not possess expert knowledge in this field, they rely on their daily experience and subconsciously find analogue situation models matching the semantic frame. Similar situations can occur in strategic games, military developments, hunting, etc. The choice depends on the background knowledge one has. This phenomenon of understanding one thing in terms of another is called *metaphor*. Our findings confirm that metaphorical mappings within the semantic frames in specialised discourse activate the background knowledge derived from similar contexts. This usefulness of embedding specialised concepts in everyday situations is highlighted by Faber/Cabezas-García (2019).

Another type of logic – intrasystem logic – sets the semiotic system's relations. It is imposed upon the real-life correspondence logic. This type of logic is described in studies carried out in grammar, morphology, syntax, etc. Our findings can also contribute to understanding the intrasystem logic, e. g. in (2) discussed earlier:

(2) Spain and Poland have been two countries traditionally targeted by *SMS scams* and similar *malware.* (GReAT 2017a)

The preposition *by* introducing the doer of the action in passive constructions works as the marker of the quasi-Agent. Meanwhile, the proposition *with* employed "for saying what is used for doing something" (The Free Online English Dictionary) is typical of the Instrument, as seen in (22):

(22) The "company" had two key "products": the malicious program, Lurk, and a huge botnet of computers infected *with it.* (Stoyanov 2016)

According to the third type of logic – communication or pragma-logic (Solomonick 2011), the same semantic frame is interpreted regarding a communicative situation in a particular discourse. For instance, the frame 'Agent *hijack* Objective' matches events in CS (23) and terrorist (24) discourses:[2]

(23) The virus […] *hijacked* another program known as Microsoft Outlook. (Christensen 1999)

(24) On Sept. 11, 2001, terrorists *hijacked* four separate planes. (Dilmore 2011)

6 Conclusion

Our research reveals the most representative SRs within thematic categories of the CS terminology. We approached this task from the cognitive perspective to rationalise the correlation between the formal SR categorisation of terms and pragmatically conditioned semantic categorisation of terms. Our findings illustrate that for interpreting CS terminology, the helpful technique is to appeal to a similar frame in another, more familiar discourse.

The classical Fillmore's frame semantic theory has been refined and extended – the quasi-Agent and Specifier SRs have been added, contributing to building a coherent system for framing events of professional communication.[3]

In our previous works (Isaeva/Burdina 2019, Isaeva/Crawford 2019), we have demonstrated the virtue of semantic framing for conceptual metaphorical modelling. The current results can be applied for cognitive mediation in professional communication to enhance specialised knowledge transfer. Based on a formalised SR distribution in a frame, it is possible to proceed to the discourse event simulation and metaphorical modelling in mental representation and cognition. Our further efforts will be fostered to apply text mining and automated metaphor identification in specialised texts.

Acknowledgements

We thank Tim and Karen Sadler (Oxford, UK) for proofreading the text.

References

Amirova, Oksana Georgievna (2002): *Semanticheskaja model anglijskih glagolov upravlenija* [Semantic Model of English Verbs of Governance]. PhD dissertation. Ufa.

Apresjan, Yurij Derenikovich (1995): *Selected Works.* 2nd ed. Vol. 1: *Lexical semantics (Synonymous means of language).* Moscow: Yazyki Russkoi Kultury.

Apresjan, Yurij Derenikovich (2009): *Issledovanija po semantike i leksikografii* [Works on Semantics and Lexicography]. Vol. 1. Moscow: Yazyki Slavyanskoi Kultury.

Brinton, Laurel J. (2000): *The Structure of Modern English: A Linguistic Introduction.* Amsterdam: Benjamins.

Chafe, Wallace L. (1970): *Meaning and Structure of Language.* Chicago: University of Chicago Press.

Cook, Walter Anthony (1998): *Case Grammar Applied.* Arlington: Summer Institute Linguistics.

[2] Examples (23) and (24) are cited from the Corpus of Contemporary American English (https://www.english-corpora.org/coca/).

[3] Given as a suggestion, which needs further investigation.

Downing, Angela / Locke, Philip (1992): *A University Course in English Grammar.* Hemel Hempstead: Prentice Hall.

Faber, Pamela / Cabezas-García, Melania (2019): "Specialised Knowledge Representation: from Terms to Frames." *Research in Language* 17.2: 197–211.

Fillmore, Charles J. (1968): "The Case for Case." *Universals in Linguistic Theory*. Eds. Emmon W. Bach / Robert Thomas Harms. New York: Holt, Rinehart & Winston. 1–88.

Fillmore, Charles J. (1971): "Types of Lexical Information." *Semantics. An Interdisciplinary in Philosophy, Linguistics and Psychology.* Eds. Danny D. Steinberg / Leon A. Jakobovits. Cambridge: Cambridge University Press. 370–392.

Fillmore, Charles J. (1982): „Frame Semantics." *Linguistics in the Morning Calm. Selected Papers from SICOL-1981*. Ed. The Linguistic Society of Korea. Seoul: Hanshin. 111–137.

Gak, Vladimir Grigorevich (1998): *Jazykovye preobrazovanija* [Language Transformations]. Moscow: Yazyki Russkoi Kultury.

Humboldt, Wilhelm von (1984): *Selected Works on Linguistics.* Moscow: Progress.

Isaeva, Ekaterina (2019): "Metaphor in Terminology: Finding Tools for Efficient Professional Communication." *Fachsprache. Journal of Professional and Scientific Communication* 41.1–2*:* 65–86.

Isaeva, Ekaterina / Baiburova, Olga / Manzhula, Oksana (2022): "Anthropomorphism in Computer Security Terminology Through the Prizm of Smart Cognitive Framing." *Science and Global Challenges of the 21st Century – Science and Technology. Perm Forum 2021*. Eds. Alvaro Rocha / Ekaterina Isaeva. Cham: Springer. 460–474.

Isaeva, Ekaterina / Burdina, Olga (2019): "Transdiscursive Term Transformation: The Evidence from Cognitive Discursive Research of the Term 'Virus'." *Current Approaches to Metaphor Analysis in Discourse*. Ed. Ignasi Navarro i Ferrando. Berlin/Boston: De Gruyter. 79–110.

Isaeva, Ekaterina / Crawford, Russ (2019): "Semantic Framing of Computer Viruses: the Study of Semantic Roles' Distribution." *Perm University Herald. Russian and Foreign Philology* 11.1: 5–13.

Katsnelson, Solomon Davidovich (2010): *Istoriko-grammaticheskie issledovanija* [Works on the history of grammar]. Saint Petersburg: Saint Petersburg Linguistic Society.

Kshirsagar, Meghana / Thomson, Sam / Schneider, Nathan / Carbonell, Jaime / Smith, Noah A. / Dyer, Chris (2015): "Frame-semantic Role Labeling with Heterogeneous Annotations." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing.* Vol. 2: *Short Papers,* Beijing. Association for Computational Linguistics. 218–224. https://aclanthology.org/P15-2036/ (06.09.2022).

L'Homme, Marie-Claude (2017): "Maintaining the Balance between Knowledge and the Lexicon in Terminology: A Methodology Based on Frame Semantics." *Lexicography* 4.1: 3–21.

Paducheva, Elena Viktorovna (2004): *Dinamicheskie modeli v semantike leksiki* [Dynamic Models in the Semantics of Vocabulary]. Moscow: Yazyki Slavyanskoi Kultury.

Plungian, Vladimir Aleksandrovich (2003): *Obshhaja morfologija: Vvedenie v problematiku* [General Morphology: an introduction]. 2nd ed. Moscow: Editorial URSS.

Rakhilina, Ekaterina Vladimirovna / Testelets, Yakov Georgievich (2016): "Nauchnoe nasledie Ch. Filmora i sovremennaja teorija jazyka" [Charles Fillmore's Legacy and Modern Theoretical Linguistics]. *Voprosy jazykoznanija* [Topics in the Study of Language] 2: 7–21.

Roth, Michael / Woodsend, Kristian (2014): "Composition of Word Representations Improves Semantic Role Labelling." *EMNLP 2014 – 2014 Conference on Empirical Methods in Natural Language Processing. Proceedings of the Conference*. 407–413. https://aclanthology.org/D14-1045/ (06.09.2022).

Saussure, Ferdinand de (1959): *Course in General Linguistics.* New York / Toronto / London: McGraw-Hili.

Solomonick, Abraham (2011): *Filosofija znakovyh sistem i jazyk* [Philosophy of Sign Systems and Language]. Moscow: LKI.

Whorf, Benjamin Lee (1956): *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, Mass.: MIT Press.

Sources of Linguistic Data

Assolini, Fabio / Makhnutin, Andrey (2013): "PAC – the Problem Auto Config." https://securelist.com/analysis/publications/57891/pac-the-problem-auto-config (10.09.2019).

Blasco, Jorge / Chen, Thomas M. (2018): "Automated Generation of Colluding Apps for Experimental Research." *Journal of Computer Virology and Hacking Techniques* 14.2: 127–138. https://link.springer.com/article/10.1007/s11416-017-0296-4 (14.09.2019).

Christensen, Damaris (1999): "Beyond Virtual Vaccinations. Developing a Digital Immune System in Bits and Bytes." *Science News* 156.5: 76–78.

Dilmore, Angie Kay (2011): "9/11 Ten Years Later." *Boys' Life*. 11 September 2011. Vol. 101 Issue 9, 26–29. Retrieved from Corpus of Contemporary American English (https://www.english-corpora.org/coca/) (06.09.2022).

"Gang of fraudsters and 'money mules' sentenced for £200k scam" (2019): *Financial Fraudster News*. https://www.financialfraudsternews.com/en/currency-fraud2/gang-of-fraudsters-and-money-mules-sentenced-for-200k-scam (19.09.2019).

GReAT (2017a): "Expensive Free Apps." *SECURELIST by Kaspersky*. https://securelist.com/expensive-free-apps/77083/ (14.09.2019).

GReAT (2017b): "Fileless Attacks against Enterprise Networks." *SECURELIST by Kaspersky*. https://securelist.com/blog/research/77403/fileless-attacks-against-enterprise-networks/ (15.09.2019).

GReAT / Naor, Ido (2016): "ATMZombie: Banking Trojan in Israeli Waters." *SECURELIST by Kaspersky*. https://securelist.com/atmzombie-banking-trojan-in-israeli-waters/73866/ (15.09.2019).

Koutlis, Manolis / Kourouniotis, Petros / Kyrimis, Kriton / Renieri, Nikolina (1998): "Inter-component Communication as a Vehicle towards End-user Modeling." *ICSE'98 Workshop on Component-Based Software Engineering, Kyoto, Japan, April 26*: 26f: https://icsa-conferences.org/series/CBSE/1998/papers/p7.html (06.09.2022).

Lad, Aditya (2011): "Top 7 Anti Rootkit Software for Windows." https://www.computerweekly.com/tip/Top-7-anti-rootkit-software-for-Windows (14.09.2019).

Makrushin, Denis (2017): "The Cost of Launching a DDoS Attack." *SECURELIST by Kaspersky*. https://securelist.com/analysis/publications/77784/the-cost-of-launching-a-ddos-attack/ (30.05.2019).

Mamedov, Orkhan / Sinitsyn, Fedor (2016): "A Malicious Pairing of Cryptor and Stealer." *SECURELIST by Kaspersky*. https://securelist.com/a-malicious-pairing-of-cryptor-and-stealer/76039/ (15.09.2019).

Marques, Thiago (2016): "The Evolution of Brazilian Malware." *SECURELIST by Kaspersky*. https://securelist.com/blog/research/74325/the-evolution-of-brazilian-malware/ (15.09.2019).

Naor, Ido / Alon, Noam (2016): "CryPy: Ransomware behind Israeli Lines." *SECURELIST by Kaspersky*. https://securelist.com/blog/research/76318/crypy-ransomware-behind-israeli-lines/ (14.09.2019).

Ortloff, Stefan (2016): "New Wave of Mirai Attacking Home Routers." *SECURELIST by Kaspersky*. https://securelist.com/new-wave-of-mirai-attacking-home-routers/76791/ (15.09.2019).

Shulmin, Alexey / Prokhorenko, Mikhail (2016): "Lurk Banker Trojan: Exclusively for Russia." *SECURELIST by Kasp*ersky. https://securelist.com/blog/research/75040/lurk-banker-trojan-exclusively-for-russia/ (15.09.2019).

Stoyanov, Ruslan (2016): "The Hunt for Lurk. How we helped to catch one of the most dangerous gangs of financial cybercriminals." *SECURELIST by Kaspersky*. https://securelist.com/analysis/publications/75944/the-hunt-for-lurk/ (15.09.2019).

"Threat Intelligence Report for the Telecommunications Industry" (2016): *SECURELIST by Kaspersky*. https://securelist.com/analysis/publications/75846/threat-intelligence-report-for-the-telecommunications-industry/ (15.09.2019).

Unuchek, Roman (2017): "Mobile Malware Evolution 2016." *SECURELIST by Kaspersky*. https://securelist.com/mobile-malware-evolution-2016/77681/ (15.09.2019).

"What is a remote server?" (2017): *Quora*. https://www.quora.com/What-is-a-remote-server (19.09.2019).

Zakorzhevsky, Vyacheslav (2015): "You can't be invulnerable, but you can be well protected". *SECURELIST by Kaspersky*. https://securelist.com/you-cant-be-invulnerable-but-you-can-be-well-protected/73160/ (15.09.2019).

Online resources and dictionaries

Cambridge Dictionaries Online. Cambridge University Press. http://dictionary.cambridge.org/ (15.09.2019).

Collins. Free Online Dictionary and Thesaurus. https://www.collinsdictionary.com/ (06.09.2022).

Computer Hope. https://www.computerhope.com/jargon.htm (15.09.2019).

English Dictionary, Thesaurus & Grammar Help (2010). https://www.lexico.com/en/ (15.09.2019).

FrameNet. https://framenet.icsi.berkeley.edu/fndrupal/ (22.08.2019).

GitHub. https://github.com/ (21.08.2019).

LexiCon Research Group (2021). http://lexicon.ugr.es/ (21.08.2021).

Techopedia. IT Dictionary for computer terms and technology (2016). https://www.techopedia.com/dictionary (15.09.2019).

Techtarget. Search Security (2021). https://www.techtarget.com/ (06.09.2022)

The Free Dictionary by Farlex: Dictionary, Encyclopedia, and Thesaurus, Collection of dictionaries by subject including medical and legal together with free and subscription encyclopedias, in ten languages (2003). http://www.thefreedictionary.com/ (14.09.2019).

The Free Online English Dictionary from Macmillan Publishers (2009). http://www.macmillandictionary.com/ (15.09.2019).

The Longman Dictionary of Contemporary English Online. http://www.ldoceonline.com/ (10.09.2019).

Your Dictionary. http://www.yourdictionary.com (15.09.2019).

Assoc. Prof. Ekaterina Isaeva, PhD
Head of the Department of English
for Professional Communication
Perm State University
15, Bukirev Str.
Perm, 614990
ekaterinaisae@psu.ru

Assoc. Prof. Oksana Manzhula, PhD
Department of English
for Professional Communication
Perm State University
15, Bukirev Str.
Perm, 614990
achilleon@mail.ru

Assoc. Prof. Olga Baiburova, PhD
Department of English
for Professional Communication
Perm State University
15, Bukirev Str.
Perm, 614990
olga3079@mail.ru

Dissertation: "I thought asking the experts of the board first before turning at the unctuous salesmen." Requesting special field related advice and information on online communities

Eveliina Salmela, Universität Vaasa, Finland

1 Introduction

Today individuals can discuss and find information effortlessly online and get peer support and advice on various issues. The focus of my research is the use of the internet as a source of peer support in situations where people need information about a topic or situation related to a specialized field. The main objective of the study is to form a comprehensive picture of how requesting and receiving special field related information function on online discussions about heat pumps in Finland. Heat pumps are devices that can be used for heating and cooling apartments and other spaces. The reason for choosing heat pumps for the topic is its specialization and popularity as a topic of discussion on online communities. I examined how a person in need of information requests information from an online community and what kind of result the request produces.

Sharing information on online communities has been studied earlier in relation to various topics, also heat pumps (Steehouder 2002, Haataja/Perttula 2004, Hyysalo/Juntunen/Freeman 2013, Hyysalo/Juntunen/Martiskainen 2018). However, in several of these studies, the primary object of interest is something other than communication. Although requesting and giving advice has been studied a lot also in the area of computer-mediated communication studies, most of the research deals with health-related topics, and there is a research gap in internet pragmatics of advice-giving related to other topics than health (Morrow 2017: 674, 681). Moreover, pragmatic research on modern online discussion boards has been largely overshadowed by other platforms in research, and a systematic and comprehensive linguistic and especially pragmatic description of discussion boards has remained incomplete (Arendholz 2017: 125, 132 f.). In my research, I fill this gap by combining a community perspective with a detailed empirical analysis that combines discursive and pragmatic examination. Peer help online has mostly been studied from the point of view of the help giver (e. g. Constant/Sproull/Kiesler 1996: 120, Wasko/Faraj 2005, Chu 2009, Fang/Chiu 2009). My research, on the contrary, examines information retrieval and sharing from the perspective of the information requester. My research utilizes data in Finnish language and belonging to the Finnish cultural context, so it also complements research with a Finnish linguistic and cultural perspective.

2 Theoretical background, data and methods

My research is located in the field of specialized communication and online communication, where it represents communication related to the special field through online peer discussions. *Technical communication* is specialized communication, where the purpose is to get the user to act in a certain way and to achieve a certain goal, which can be, for example, assembling, installing or using a product (Walter 1996: 30). In this study, I examine technical communication, which takes place as user peer communication on a voluntary basis. From a

technical communication perspective, discussion boards can help users solve problems that they cannot solve with user documentation (Steehouder 2002: 491).

The data of the dissertation consists of 300 message threads about heat pumps that were manually collected from two Finnish discussion boards, Lampopumput.info (heatpumps.info) and Suomi24.fi (Finland24.fi). The total number of messages examined was more than six thousand (6,078). I chose two platforms that are central to the heat pump-related discussion and are popular, but different in nature. Lampopumput is Finland's largest discussion forum related to heat pumps. Suomi24 instead is a central platform in Finnish online discussions because of its long history and large number of users (e. g. Harju 2018: 59, 70). Lampopumput is a specialized forum for special field enthusiasts and experts, and Suomi24, on the other hand, is a general forum for almost all discussion topics. Lampopumput requires authors to register for the site, while on the Suomi24 forum it is possible to write messages completely anonymously. This offers the opportunity to examine what kind of connection the anonymous participation option has with how users communicate about the topic.

The methodological framework of the study is built on computer-mediated discourse analysis (CDMA) and pragmatics of computer-mediated communication. The subject of the analysis are the texts available online. CDMA can include the analysis of all kinds of online behavior based on empirical, textual observation (Herring 2004: 2). In this study, I use language-focused content analysis as part of CMDA (Herring 2004: 2–5). I use both quantitative and qualitative research methods. The goal is to create an overview of the phenomenon under study using quantitative methods and to deepen the understanding using qualitative methods. Non-randomness of the results was ensured by checking the results using statistical methods.

The classification framework of the data was formed in a dialogue between theory and data. In the first research phase, the opening messages were classified into different discussion themes and according to their communicative goal. The second research phase focused on those message threads (N=195) whose opening message aimed at obtaining information. I analyzed the structure of the messages, the means of expression, such as multimodal means (e. g. emoticons), politeness and humor used in the messages. I also looked at the number and distribution of responses and what kind of connection the various discussion themes and goals and other characteristics of the opening messages had with the number of responses they received. In the third research phase, I examined the entire message threads, whose opening message aimed at getting information. For this purpose, I used a smaller random sample of the data (40 message threads with 764 answers, 382 from both discussion boards). I classified the opening messages according to the directness of the information request, specificity of the problem or question raised in the message, and emotionality of the message. Finally, I looked at what kind of argumentation was used in the reply messages.

3 Findings

Based on the analysis, the main topics of discussion were the purchase process of products and services, installation and use of products, and general technical discussion. Of these, the purchase process of products and services, was the most discussed, and use of devices the second most discussed topic. In the open discussion board, opening messages related to a purchase decision and buying products were clearly more common than on the controlled discussion board. Open discussion board which offers the possibility to write anonymously may offer a lower threshold for discussing issues related to the purchase situation (cf. Savolainen 2001:

76–78). The user pool is also likely more heterogeneous and the forum can be more approachable than the specialized discussion board.

The main goals of the opening messages were getting information, giving information and stimulating discussion. Getting information was the most common goal of opening messages in the data. This is in line with previous studies (e. g. Arendholz 2013, Bakke 2019) which have found that the most common reason for starting a new message thread is to ask questions for information or advice. Opening messages that aimed at obtaining information were divided into requests for advice, factual questions, and requests for opinions, suggestions and experiences. The most common of these was a request for advice, which accounted for a fifth of all messages. Asking for advice and offering information were more common on the controlled discussion board, while expressing an opinion and asking for experiences were more common on the open discussion board.

The basic structural elements of the opening messages were information request, problem description, technical information, evaluation/justification and plan/proposed solution. All of these belong to an informational-expository schema, the transitions of which present and develop a solution to the problem (cf. Herring 1996: 85). In addition to these, the opening messages included interactive structural elements, which include greeting, preface, metacomment, showing activity, appeal to readers, politeness and signature. In my study I found, that opening messages with a greeting, compliment or meta-comment received more responses than on an average. All in all, the authors can use the optional structural elements to show that they respect and value the forum, the message readers, the respondents and their expertise and do not use the forum just as an "answering machine".

Requesting information was most often direct in the opening messages, although it is generally considered as greater threat to face than indirect requests. Expression of feelings was rare in opening messages. If feelings were described, they described frustration and irritation towards, for example, non-functioning technology. The most common multimodal means of communication in the data were emoticons and hyperlinks. Emoticons were used in about every tenth opening messages. Emoticons can strengthen positive speech acts or soften face-threatening speech acts, for example, requests (Derks et al. 2008: 386, Skovholt/Grønning/Kankaanranta 2014: 780). Hyperlinks used to refer to information sources relevant to the topic (cf. Stommel/Paulus/Atkins 2017: 59–61).

The use of humor in the opening messages was mainly reinforcing, i. e. building consensus and being polite. Only a few messages used subversive, questioning humor, and humor attacking other groups was not used at all. Humor types and methods used in reinforcing humor were funny figures of speech, idioms, personification of technology, self-irony and self-deprecation. Humor was often based on shared knowledge or experiences. The basis could be general cultural knowledge, knowledge related to a specialized field, or knowledge related to the discussion board's previous activity. The humor was often linked to the special field, so understanding it required some knowledge of the topic under discussion, and the goals of the humor also arose from the topic of the discussion. Using reinforcing humor shows goodwill and politeness towards the readers of the message, helps to build a common ground and to create a consensus. This is understandable when the goal is to get help or information, in which case giving a pleasant, sympathetic image of oneself is important. Humor was also used to some extent to emphasize one's own point of view. Subversive humor was used to question and criticize things. With the help of humor, a negative or critical message could be conveyed in a more acceptable form. For example, similes and irony were

used as types of subversive humor. On the controlled discussion board, the use of humor in opening messages aimed at obtaining information was more common than on the open discussion board.

Almost 40 % of the answer messages in the data were irrelevant. On the open discussion board, a larger proportion of the answers, even more than half, were irrelevant, while on the controlled forum only 22.5 %. The relevant reply messages were divided into advice, giving general information, sharing one's own experiences, opinions and questions. The most common of these was sharing one's own experiences, and almost as common was giving advice. Almost a third of the relevant answers in both discussion boards were descriptions of the writer's own experiences.

The writers used facts, own experiences, personal beliefs, other people's opinions and emotive appeals to support their claims on the answer messages. Personal beliefs were the most frequently used justification for claims on both discussion boards. The results are in line with previous studies which indicate that in digital communication environments, citizens' own experiences may be more convincing than the research-based justification presented by an expert in many situations, and they are also a very commonly used means of justifying claims in online discussions (Lehti/Kallio 2017, Lehti/Eronen-Valli 2018). On the specialized, controlled discussion board, the writers used more often factual information based on some information source and online information sources to justify the claims when compared to the open discussion board, where other people were used as a source of information more often. This indicates that on a controlled discussion board, people pay more attention to the justification of their claims and thereby also to their own reputation as a provider of reliable information.

A fixed online identity creates a pressure to develop this identity and function as a member of the online community. Membership of the community also brings with it obligations, for example, reciprocity and activity in the community: that you also give something to the community from which you have received help. In the open and anonymous discussions, on the other hand, there is more room for playful and entertaining, but also for sharp critical discussion, when the authors do not have to worry so much about personal sanctions.

4 Conclusion

Search for information online and how practical everyday technological problems are discussed is often a socially complex situation, the mechanisms of which this study sheds light on. My research contributes new information about requesting information and giving advice, and the appropriate form of the message in terms of achieving the goal of the communicator. In terms of digital communication and digital discourse analysis, my research brings new information about the digital discussion cultures and the role of anonymity in them. Compared to previous studies related to the same topics (cf. e. g. Steehouder 2002, Locher 2006, Kouper 2010, Arendholz 2013, Haasio 2015, Hyysalo/Juntunen/Martiskainen 2018), my research is multidimensional in nature: the research includes several perspectives, while previous studies have typically focused on one for examining a narrower perspective. A comparison of open and a controlled discussion boards has not been done before in a similar context. The study succeeds in describing the differences in two different types of data collected from two different sites in relation to many different factors. With my study, I have been able to bring out features and details from the both types of data that previous research has not been able to reach.

The reason for this is the perspective chosen in the study and research methods that have not been applied to same kind of data before.

The research offers both empirical and theoretical contribution to the research of specialized communication, digital discourse and interaction, as well as information acquisition. My research contributes a methodological and analytical framework, which can be used for other types of materials in the future. The results can also be applied to the research of information search and sharing via online communities in other subject areas, especially technical specialized fields, as well through other platforms than discussion boards, taking place in text-based digital interaction. Conducting comparative research, for example, on subjects related more purely to professions, hobbies or everyday life problems could create interesting aspects for further research. Other different types of platforms to which the classification of message types could be applied are, for example, WhatsApp and Facebook groups and Jodel.

References

Arendholz, Jenny (2013): *(In)Appropriate Online Behavior: A Pragmatic Analysis of Message Board Relations.* Amsterdam: Benjamins.

Arendholz, Jenny (2017): "Message Boards." *Pragmatics of Social Media.* Eds. Christian Hoffmann / Wolfram Bublitz. (Handbooks of Pragmatics [HOPS] 11). De Gruyter Mouton. 125–150.

Bakke, Abigail (2019): "Trust-building in a Patient Forum: The Interplay of Professional and Personal Expertise." *Journal of Technical Writing and Communication* 49.2: 156–182.

Chu, Kuo Ming (2009): "A Study of Members' Helping Behaviors in Online Community." *Internet Research* 19.3: 279–292.

Constant, David / Sproull, Lee / Kiesler, Sara (1996): "The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. *Organization Science* 7.2: 119–135.

Derks, Daantje / Bos, Arjan / Grumbkow, Jasper von / Von, Jayce (2008): "Emoticons and Online Message Interpretation." *Social Science Computer Review* 26.3: 379–388.

Fang, Yu-Hui / Chiu, Chao-Min (2009): "In Justice we Trust: Exploring Knowledge-sharing Continuance Intentions in Virtual Communities of Practice." *Computers in Human Behavior* 26: 235–246.

Haasio, Ari (2015): *Toiseus, tiedontarpeet ja tiedon jakaminen tietoverkon "pienessä maailmassa: Tutkimus sosiaalisesti vetäytyneiden henkilöiden informaatiokäyttäytymisestä*. (Acta Universitatis Tamperensis 2082). Tampere: Tampere University Press.

Haataja, Rauli / Perttula, Juha (2004): "Asiantuntijuuden tuottaminen nuorten Internetkeskustelualueilla." *Psykologia* 39.5: 355–369.

Harju, Auli (2018): "Suomi24-keskustelut kohtaamisten ja törmäysten tilana." *Media & Viestintä* 41.1: 51–74.

Herring, Susan C. (1996): "Two Variants of an Electronic Message Schema." *Computer-Mediated-Communication. Linguistic, Social and Cross-Cultural Perspectives.* Ed. Susan C. Herring. Amsterdam/Philadelphia: Benjamins. 81–106.

Herring, Susan C. (2004): "Computer-mediated Discourse Analysis. An Approach to Researching Online Behavior." *Designing for Virtual Communities in the Service of Learning*. Eds. Sascha A. Barab / Rob Kling / James H. Gray. Cambridge: Cambridge University Press. 338–376.

Hyysalo, Sampsa / Juntunen, Jouni K. / Freeman, Stephanie (2013): "Internet Forums and the Rise of the Inventive Energy User." *Science & Technology Studies* 26.1: 25–51.

Hyysalo, Sampsa / Juntunen, Jouni K. / Martiskainen, Mari (2018): "Energy Internet Forums as Acceleration Phase Transition Intermediaries." *Research Policy* 47.5: 872–885.

Kouper, Inna (2010): "The Pragmatics of Peer Advice in a LiveJournal Community." *Language@Internet* 7: article 1. https://www.languageatinternet.org/articles/2010/2464 (05.09.2022).

Lehti, Lotta / Eronen-Valli, Maria (2018): "Diskurssintutkimuksen menetelmiä digitaalisen retoriikan tutkimuksessa." *Diskurssintutkimuksen menetelmistä. On the methods in discourse studies.* Eds. Lauri Haapanen / Liisa Kääntä / Lotta Lehti. (AFinLA-e Soveltavan kielitieteen tutkimuksia 11). Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. 156–176.

Lehti, Lotta / Kallio, Johanna (2017): "Participation in an Online Social Policy Discussion: Arguments in Focus." *Discourse, Context & Media* 19: 58–65.

Locher, Miriam A. (2006): *Advice Online: Advice-giving in an American Internet Health Column.* Amsterdam, NLD: Benjamins.

Morrow, Phillip R. (2017): "Requesting and Advice-giving." *Pragmatics of Social Media.* Eds. Christian Hoffmann / Wolfram Bublitz. Berlin/Boston: De Gruyter Mouton. 661–689.

Savolainen, Reijo (2001): "Living Encyclopedia or Idle Talk? Seeking and Providing Consumer Information in an Internet Newsgroup." *Library and Information Science Research* 23.1: 67–90.

Skovholt, Karianne / Grønning, Anette / Kankaanranta, Anne (2014): "The Communicative Functions of Emoticons in Workplace E-mails:-)*." *Journal of Computer-Mediated Communication* 19.4: 780–797.

Steehouder, Michaël F. (2002): "Beyond Technical Communication: Users Helping each other." *Proceedings. IEEE International Professional Communication Conference 2002*: 489–499.

Stommel, Wyke / Paulus, Trena M. / Atkins, David P. (2017): "'Here's the Link': Hyperlinking in Service-focused Chat Interaction." *Journal of Pragmatics* 115: 56–67.

Wasko McLure, Molly / Faraj, Samer (2005): "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice." *MIS Quarterly* 29.1: 35–57.

Walter, John A. (1996): "Technical Writing: Species or Genus?" *Defining Technical Communication*. Ed. Dan Jones. Arlington: Society for Technical Communication, Inc. 27–30.

FACHSPRACHE Journal of Professional and Scientific Communication

FACHSPRACHE – Journal of Professional and Scientific Communication is a refereed international journal that publishes original articles on all aspects of specialized communication and provides an interdisciplinary forum for researchers and teachers interested in this field. **FACHSPRACHE** is committed to promoting high-quality research and to improving professional and scientific communication and knowledge transfer including the qualifications required for this. The journal places equal emphasis on rigour and comprehensibility in an attempt to bridge the gap between researchers who may come from different disciplines but who share a common interest in professional and scientific communication. These disciplines may range from linguistics, psychology, lexicography, terminology, and translation studies to the neurosciences as well as domains such as business administration, economics, law, medicine, science and engineering. Special issues focusing on particular individual areas are published regularly.

ISSN 1017-3285 **https://ejournals.facultas.at/index.php/fachsprache**