

Faculté de philosophie, arts et lettres

Collocations across Registers

Auteur : Lin Ding

Promoteur : Magali Paquot

Année académique 2025-2026 – September session

LING2MS/LM Master [120] in Linguistics, with a specialized focus: Linguistics and its Applications for a Multilingual Society

Declaration of honour

Dissertation/Individual Final Project

Faculty of Philosophy, Arts and Letters, UCLouvain

I hereby declare that this is an original and personal work, that all referenced sources have been fully cited, regardless of their origin, and that the use of artificial intelligence tools complies with the general instructions published by UCLouvain and available at the following address :

<https://cdn.uclouvain.be/groups/cms-editors-p1/portail/reglement/Consignes-ChatGPT-version-etudiante.pdf?itok=SebXXm87> (French version)

I am aware that failure to cite a source, to cite it clearly and fully, to announce in a specific section at the beginning of the dissertation the use I have made of artificial intelligence tools, or to use them in a manner contrary to the guidelines set forth in the institutional note, constitutes serious irregularities in accordance with the [Academic regulations and procedures](#) (Chapter 4, section 7, articles 107 to 114) within the University. In particular, I am aware of the risks of academic and disciplinary sanctions.

In light of the above, I declare on my honour that I have not committed plagiarism or any other form of fraud and that I have not used artificial intelligence tools beyond what is permitted by UCLouvain.

Last name, First name : Lin Ding

Date : 16 Aug, 2026

Student's signature : *Lin Ding*

Acknowledgements

I am deeply grateful to Professor Magali Paquot for supervising this dissertation and for giving me the opportunity to pursue a topic in which I have a strong personal interest. I would also like to thank her for taking the time to read and comment on multiple versions of my research proposal and chapter drafts. Her patience, guidance, constructive feedback, and continued support throughout the research and writing process have been invaluable. This dissertation would not have been possible without her encouragement and support.

Abstract

This dissertation investigates the use of 25 high-frequency lexical verbs shared across informal conversation, parliamentary debates, news reports, and academic prose. Using manually validated relational collocations extracted from Sketch Engine, it examines differences in the grammatical relations in which these verbs occur, the frequency and distribution of their relational collocations, their degree of cross-register sharing, and the semantic patterns of collocations within the object relation. The results show clear register variation in the grammatical and collocational patterns of the 25 lexical verbs, and this variation is both verb-specific and relation-specific. These findings provide insights into register-specific language use and have implications for second language acquisition and teaching.

Keywords: register variation; high-frequency lexical verbs; relational collocation; corpus-based analysis

Table of Contents

1	Introduction	1
2	Literature Review	3
2.1	Introduction	3
2.2	Phraseology	3
2.2.1	Definition and approaches to phraseology	3
2.2.2	Scope and types of phraseological units	5
2.2.3	Collocations as core phraseological units	6
2.2.3.1	Conceptualisation and operationalisation of collocations	6
2.2.3.2	Quantitative measures in collocation research	7
2.2.3.3	Methodological issues in collocation research	8
2.2.4	The role of collocations in language use and research	10
2.2.4.1	The role of collocations in meaning construction	11
2.2.4.2	The role of collocations in L2 learner performance	11
2.2.4.3	Collocational resources and analytical approaches	13
2.3	Register	15
2.3.1	Register and its analytical framework	15
2.3.2	Conceptualising and analysing register variation	24
2.3.2.1	Empirical findings of studies using the individual feature approach	25
2.3.2.2	Empirical findings of studies using MD analysis	26
2.4	Phraseological patterning across registers: recurrence and co-occurrence	30
2.4.1	Recurrence: lexical bundles	30
2.4.2	Co-occurrence: statistical collocations	33
2.5	Positioning of the present study	37
3	Methodology	38
3.1	Introduction	38
3.2	Corpora	38
3.3	Selection of target lexical verbs	40
3.4	Extraction and validation of relational collocations	40
3.5	Analytical procedures	44
3.5.1	Analysis of register variation in verb collocations	44
3.5.2	Analysis of frequency, distribution, and register sharing	45
3.5.3	Semantic and functional interpretation of collocational patterns	45
3.5.4	Descriptive and inferential statistics	47
3.6	Conclusion	48
4	Results	49
4.1	Introduction	49
4.2	Grammatical Relations across the 25 Verbs: Distribution and Register-Distinctive Patterns	49
4.2.1	Distribution of attested grammatical relations	49
4.2.2	Register-distinctive patterns across the 25 verbs	52
4.3	Case Study: Grammatical and Collocational Patterns of <i>Say</i>	54
4.3.1	Grammatical behaviour of <i>say</i> across registers	54

4.3.2	Evaluation of tagged data	56
4.3.3	Distribution of relational collocations for <i>say</i> across registers and six selected grammatical relations	58
4.3.4	Shared and register-specific relational collocations of <i>say</i>	60
4.3.5	Interim summary	69
4.4	Distribution of Relational Collocations across Registers and Three Selected Grammatical Relations for the 25 Verbs	70
4.4.1	Introduction	70
4.4.2	Overall distribution of relational collocations across registers and the three grammatical relations	71
4.4.3	Distribution of relational collocations by grammatical relation and register	72
4.5	Shared and Register-Specific Relational Collocations across the 25 Verbs	79
4.5.1	Introduction	79
4.5.2	Overall distribution of shared and register-specific collocations	79
4.5.3	Relational collocations shared across all four registers	80
4.5.4	Relational collocations shared across three registers	82
4.5.5	Relational collocations shared across two registers	84
4.5.6	Register-specific relational collocations	86
4.5.7	Semantic analysis of relational collocations in the object relation across registers for the 25 verbs	89
4.6	Cross-verb synthesis and summary	93
5	Discussion	98
5.1	Introduction	98
5.2	Overall register variation in relational collocations	98
5.3	The influence of register on the frequency and distribution	101
5.4	Collocational patterns, communicative purposes, and discourse functions	102
5.4.1	Introduction	102
5.4.2	Semantic patterns of shared object collocates	103
5.4.3	Semantic patterns of register-specific object collocates	107
5.4.4	Summary of semantic patterns across registers	110
5.5	Conclusion	111
6	Conclusion	112
	References	114

1 Introduction

Lexical verbs, primary verbs (*be*, *have*, and *do*), and modal verbs constitute the three major classes of verbs in English. Lexical verbs form an open class whose members function as the head of the verb phrase and carry its main semantic content (Biber et al., 2021, p. 356). Within this class, a relatively small set of basic lexical verbs consistently ranks highly in corpus-derived frequency lists. These high-frequency lexical verbs are important not only because of their frequency but also because they have broad, often polysemous meanings and occur across a wide range of communicative contexts. Altenberg and Granger (2001) relate this polysemy to two kinds of meaning extension: a universal tendency towards “more general, abstract, delexicalized or grammaticalized uses” and various language-specific tendencies that give rise to “specialized meanings, collocations, and idiomatic uses” (p. 174). As a result, the meanings and uses of these verbs depend heavily on the lexical and grammatical environments in which they occur. For example, a verb such as *make* can take on different meanings in combinations such as *make a decision* and *make money*. This context-dependent nature makes high-frequency lexical verbs particularly relevant to the study of verb collocations.

Given this sensitivity to context, examining verb collocations from a register perspective is particularly important. Corpus-based register studies have shown that linguistic features vary across registers at lexical, grammatical, and lexico-grammatical levels (e.g., Biber, 2012). Such variation is associated with differences in the situational characteristics of communicative contexts, including participants, participant relationships, mode, processing circumstances, setting, communicative purposes, and topic (Biber & Conrad, 2019). Since these contextual factors shape language use, register may influence not only the frequency and distribution of individual lexical verbs, but also the collocates with which they combine and the grammatical relations in which these combinations occur.

Evidence from previous research also suggests that lexical verbs tend to occur in preferred lexico-grammatical patterns within particular communicative contexts. For example, Granger and Paquot’s (2009) study of lexical verbs in English for Academic Purposes (EAP) shows that individual verbs in academic discourse have their own “preferred lexico-grammatical company”, including preferred subjects (*this study shows that; the evidence suggests that*), objects (*support the view/hypothesis that; provide evidence/information*), adverbs (*differ significantly; vary considerably/widely*), and routinized structures (*as discussed in; it should be noted that*) (p. 3). However, existing studies of high-frequency verbs often focus on a single register or disciplinary domain and therefore provide limited evidence of how such patterns vary across communicative contexts.

A broader gap concerns research on phraseological patterning across registers. Previous studies have paid more attention to recurrence-based patterns, such as lexical bundles and other recurrent contiguous sequences, than to co-occurrence-based patterns, such as statistical collocations. Further research is thus needed to clarify how register shapes verb collocations as part of phraseological patterning.

The present study addresses these gaps by examining variation in verb-based collocational patterns across four registers: informal conversation, parliamentary debates, news articles, and academic journal articles. It focuses on 25 high-frequency lexical verbs shared across these registers and analyses their relational collocations, understood here as statistically associated verb-collocate combinations identified within specific grammatical relations. The four registers are selected because they present both contrasts and continuities. Informal conversation and academic journal articles represent markedly different communicative settings, while parliamentary debates and news articles share an informational and argumentative orientation and occupy the more

formal end of the continuum. The last two registers differ in medium: parliamentary debates are spoken (often written-to-be-spoken), whereas news texts are written. A further distinction concerns their overall discourse orientation: parliamentary debates are more directive and are oriented towards influencing the course of events, whereas news discourse is more reflective and concerned with interpreting events, although both tendencies may be present in each register (Granger, 2014). Together, the four registers allows for an examination of verb collocations across different communicative settings and along the spoken–written continuum.

In light of these considerations, the present study examines verb collocations from a register-based perspective, with particular attention to how their patterns vary across communicative contexts. Collocations are analysed as lexico-grammatical patterns that may contribute to register differentiation, while register is examined as the communicative context in which these patterns vary in frequency, distribution, semantic preference, and discourse function. By bringing together insights from phraseology, corpus linguistics, and register studies, the study aims to provide a more integrated account of verb collocations across registers and to show how such variation relates to the communicative purposes and discourse functions of individual registers.

This dissertation is structured as follows. Chapter 2 reviews previous research on phraseology and collocation, register and register variation, and phraseological patterning across registers. Chapter 3 describes the corpora, the selection of target lexical verbs, the extraction and validation of relational collocations, and the analytical procedures. Chapter 4 presents the results of the corpus analysis, and Chapter 5 discusses these results in relation to the research questions. Chapter 6 concludes the dissertation by summarising the main findings and outlining the study's limitations.

2 Literature Review

2.1 Introduction

This chapter reviews the theoretical and empirical foundations underpinning the present study, with a focus on the interaction between phraseology and register variation. **Section 2.2** introduces phraseology as a field, outlining its key definitions and approaches, as well as the scope and types of phraseological units, before narrowing the focus to collocations as core phraseological units. It further discusses the conceptualisation and operationalisation of collocations, the quantitative measures used in collocation research, and related methodological issues. The section also highlights the role of collocations in language use and research.

Section 2.3 shifts to the notion of register, presenting its conceptualisation and analytical frameworks and reviewing approaches to register variation, including both individual feature analysis and multidimensional analysis. **Section 2.4** brings these strands together by examining research on phraseological patterning across registers in terms of recurrence and co-occurrence. It first considers lexical bundles, especially verb phrase-based bundles, and their functions as register discriminators, alongside methodological limitations. It then turns to statistical collocations, focusing on variation in frequency and association strength, and reviews selected verb-based case studies. Finally, **Section 2.5** identifies research gaps and outlines research questions, thereby positioning the present study within the existing literature.

2.2 Phraseology

2.2.1 Definition and approaches to phraseology

From being a peripheral subfield of lexicology, phraseology—“the study of the structure, meaning and use of word combinations” (Cowie, 1994, p. 3168)—now occupies a much more central position in language studies, largely propelled by advances in corpus linguistics (Granger, 2021). A growing body of research within corpus linguistics is now dedicated to investigating “phraseological items, patterns, constructions, or multi-word units” (Römer, 2009, p. 141). As a result, phraseology has become a key concept in both theoretical linguistics (e.g., construction grammar) and applied linguistic research, with corpus linguistics serving as the dominant methodological framework for its investigation (Gries, 2008).

The predominance of corpus linguistics methods is not accidental. On the one hand, advances in corpus compilation and large-scale corpus analysis have facilitated the systematic exploration of patterns of language use in the linguistic output of both first-language (L1) and second-language (L2) speakers. These empirically identified patterns provide a robust basis for formulating and testing hypotheses about language learning, processing and use (Gablasova et al., 2017). On the other hand, one of the most influential findings to emerge from corpus-based research is that “collocations and other types of multi-word combinations are ubiquitous in natural language” (Forsberg Lundell, 2021, p. 370). Studies report that formulaic sequences account for approximately 30%–50% of language use overall (Conklin & Schmitt, 2012; Vilkaitė, 2016). This insight has revitalized lexical studies and prompted what Granger (2021) describes as “a complete overhaul of the theory and practice of phraseology” (p. 5). It has also contributed to a gradual shift away from a traditional single-word focus towards the recognition of multiword units as a central component of linguistic competence and performance (Granger, 2018).

A similar trend is observed by Gray and Biber (2015), who note that since the 1990s most research on phraseological patterns has been empirical in nature, making extensive use of corpus analysis. In this line of research, two main issues have remained central: “the best way to define

and identify fixed multi-word units, and analysis of the discourse functions that these multiple word units perform” (Gray & Biber, 2015, p. 125). With respect to the first issue, the two main approaches typically recognized are the *phraseological approach* and the *distributional* (or *frequency-based*) *approach*. The phraseological approach seeks to establish linguistic criteria for distinguishing different types of word combinations, often focusing on properties such as semantic non-compositionality, lexical restriction, and syntactic fixedness. In comparison, the distributional approach relies primarily on quantitative criteria, identifying word combinations on the basis of their frequency and statistical association in corpus data (Granger & Paquot, 2008; Paquot & Granger, 2012). By adopting this distributional perspective, the scope of phraseology has been considerably broadened, positioning it at the intersection of four major linguistic domains: semantics, morphology, syntax, and discourse (Granger & Paquot, 2008).

With regard to semantics, semantic co-occurrence restrictions and the notion of non-compositionality play a central role in the identification and classification of word combinations in the traditional phraseological approach. In contrast, the distributional approach to multi-word units is grounded in the “Firthian contextual theory of meaning” (Granger & Paquot, 2008, p. 4), according to which meaning is not confined to individual words but emerges from their contextual environment. Within this framework, the systematic relationship between a word and its environment is manifested in two forms of meaning: *semantic preference* and *semantic prosody* (discussed later in Section 2.2.4.1). These notions have contributed significantly to the theoretical accounts of how multi-word units are formed (Granger & Paquot, 2008).

Phraseology also interacts with morphology, largely due to the ambiguous definition of the concept of *word*. In particular, the treatment of (solid, hyphenated, open) compounds and categories (e.g., complex prepositions, adverbs and conjunctions) together raises classification issues, which blur the boundary between single words and multi-word units and therefore affect the scope of phraseology (Granger & Paquot, 2008).

A similar blurred boundary can be observed between phraseology and syntax. This is particularly evident in the case of collocations, especially the notion of grammatical collocations (i.e., combinations of a lexical word with a grammatical word or structure). Another area of overlap concerns the distinction between compounds and syntactic phrases, as some word combinations “straddle the lexicon-syntax divide” (Granger & Paquot, 2008, p. 6). In addition, the degree of syntactic flexibility has often been considered an indicator of phraseological or idiom status. In this respect, while the traditional approach tends to maintain a clear distinction between phraseology and syntax, the distributional perspective places greater emphasis on the “lexico-grammar interface” (Granger & Paquot, 2008, p. 6).

Phraseology is also closely intertwined with discourse. The traditional approach mainly emphasizes pragmatics, favouring “units that reflect discourse as interaction”. In contrast, the distributional approach shifts the focus towards stylistics or rhetoric, giving more attention to “text structuring multi-word units” (Granger & Paquot, 2008, p. 7).

The intersections between phraseology and semantics, morphology, syntax, and discourse therefore highlight its inherently multidisciplinary nature, which has inevitably resulted in the vast and confusing terminology used to describe different types of word combinations. Across different research traditions, such combinations have been examined under various labels, including *lexical bundles* in corpus linguistics and register studies, *n-grams* in natural language processing, and *formulaic language* or *formulaic sequences* in second language acquisition and psycholinguistics. Despite these terminological differences, these strands of research converge on a shared concern with “how words combine into more and less fixed combinations” (Gray & Biber, 2015, p. 125). Consequently, the issues of “identifying and classifying phraseologisms and integrating them into theoretical research and practical application” have profoundly influenced

researchers across different sub-disciplines of linguistics as well as related areas such as language learning, acquisition, and teaching (Gries, 2008, p. 3).

2.2.2 Scope and types of phraseological units

With a view to clarifying the defining dimensions of phraseological phenomena and making the overlap in assumptions, concepts and findings across different research traditions more transparent, Gries (2008) proposed six criteria for the rigorous definition of co-occurrence phenomena:

- i. the nature of the elements involved in a phraseologism;
- ii. the number of elements involved in a phraseologism;
- iii. the number of times an expression must be observed before it counts as a phraseologism;
- iv. the permissible distance between the elements involved in a phraseologism;
- v. the degree of lexical and syntactic flexibility of the elements involved;
- vi. the role that semantic unity and semantic non-compositionality/non-predictability play in the definition. (p. 4)

Based on these parameters, a *phraseological unit* or *phraseologism* can be defined as “the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance” (Gries, 2008, p. 6).

Given this broad definition, the range of linguistic phenomena regarded as phraseological units is very large. It includes various types of word combinations, such as fixed expressions (e.g., *to run amok*), phrasal verbs (e.g., *to pick up*), and constructions (e.g., the *into*-causative construction). In this sense, phraseological units can serve as “a convenient cover term for co-occurrence phenomena at the syntax-lexis interface” (Gries, 2008, p. 8). Therefore, different types of phraseological units can be placed on a cline ranging from variable, transparent combinations to fixed, opaque expressions, depending on their degree of semantic non-compositionality, syntactic fixedness, lexical restriction, and institutionalization (Granger & Paquot, 2008).

From a corpus-based distributional perspective, these phraseological units can be approached in terms of two broad forms of phraseological patterning: *recurrence* and *co-occurrence*. The former is typically investigated through *n-gram analysis*, while the latter is identified through *co-occurrence analysis*. **N-gram analysis** focuses on “the extraction of recurrent continuous sequences of two or more words”, typically relying on relatively high frequency thresholds (Granger & Paquot, 2008, pp. 10-11). In this context, recurrence refers to “the repetition of contiguous strings of words of a given length (e.g., *bigrams* or *trigrams*)”. Such recurrent strings are commonly called *n-grams*, *lexical bundles*, and *clusters* (Paquot & Granger, 2012, p. 138). Lexical bundles, for example, are “fully corpus-driven sequences of contiguous words which recur frequently in speech and writing”, and often fail to “display syntactic integrity” despite functioning as “building blocks that characterize fluent spoken and written discourse” (Granger, 2018, pp. 235-236).

Co-occurrence analysis, by contrast, allows for “the statistical uncovering of significant word co-occurrences” (Granger & Paquot, 2008, p. 10). Co-occurrence involves “the co-selection of (usually) two lexical items, which may be, but are not necessarily, contiguous” (Paquot & Granger, 2012, p. 136). The retrieved units are referred to as *co-occurrences* (or *co-occurents*) rather than the traditional terms *collocations* or *collocates* (Paquot & Granger, 2012). The precise linguistic units identified through co-occurrence analysis depend on how the relationship between co-occurring words is defined and operationalized. Since collocations constitute the main focus of the present study, the following section examines their conceptualisation and operationalisation in corpus-based research in greater detail.

2.2.3 Collocations as core phraseological units

2.2.3.1 Conceptualisation and operationalisation of collocations

Since the term *collocation* was first coined by J. R. Firth in the 1930s, it has been interpreted and operationalised in various ways across studies. This lack of consensus has resulted in considerable terminological ambiguity and has complicated comparison across studies (Xiao, 2015). The present study therefore follows Evert (2005) in treating *collocation* as a generic term whose specific meaning can be narrowed down depending on particular research questions and extraction methods.

At the theoretical level, two broad conceptualisations are distinguished: *restricted collocations*, associated with the traditional phraseological approach, and *statistical collocations*, associated with the distributional or frequency-based approach. In the traditional view, **restricted collocations** are defined as “pairs of words that are in a syntactic relation, and display restricted commutability and some degree of semantic opacity or specialization” (Granger, 2018, p. 230). By contrast, from a frequency-based perspective, **statistical collocations** are discontinuous word pairs that “co-occur more often than their respective frequencies and the length of text in which they appear would predict” (Jones & Sinclair, 1974, p. 19).

While restricted and statistical collocations differ primarily in how collocation itself is conceptualised, corpus-based research also varies in how co-occurrence between two words, or *word pairs*, is defined. Evert (2005) draws a broad distinction between *positional* and *relational co-occurrences*. **Positional co-occurrences** are defined in terms of textual proximity, either within a specified collocational span or within linguistically motivated windows such as clauses, sentences, paragraphs, or documents. **Relational co-occurrences**, by contrast, are based on “a linguistic interpretation of the observable corpus data”, with each co-occurrence corresponding to “an instance of a specific structural relation”. Such relations may include graphemic adjacency, dependency relations, and relations established through phrase-structure analysis (pp. 18-19).

In a later account, Evert (2008) distinguishes three approaches to defining co-occurrence: *surface*, *textual*, and *syntactic co-occurrence*. **Surface co-occurrence** is based on a predefined collocational span, typically measured by “the number of intervening word tokens”, so that collocates are identified “within the collocational spans around the instances of a given node word” (pp. 1220-1221). A limitation of this approach is that the choice of span size may be arbitrary. **Textual co-occurrence**, by contrast, is defined by occurrence of two words within the same textual unit, such as a sentence or document, irrespective of their linear distance. Both surface co-occurrence and textual co-occurrence can therefore correspond closely to the positional approach described by Evert (2005). **Syntactic co-occurrence** is more restrictive, requiring the co-occurring words to be connected by a direct or indirect syntactic relation. It thus corresponds most closely to the relational model of co-occurrences in Evert’s (2005) framework.

A limitation of positional co-occurrence, however, is that it does not capture syntactic relations such as subject–verb, verb–object, predicative adjective, and verb–particle. Some of these relations have been reported to be particularly challenging for EFL learners, even at an intermediate to advanced level (Paquot et al., 2020). Syntactic co-occurrence, in contrast, is particularly useful when there are long-distance dependencies between collocates, since it is not constrained by an arbitrary span and introduces less noise than textual co-occurrence. It is also well suited to multiword extraction, as many lexicalised multiword expressions occur in specific syntactic patterns (Evert, 2008). For these reasons, the present study adopts a statistical conception of collocation and a relational model of co-occurrence, operationalised through syntactic relations.

2.2.3.2 Quantitative measures in collocation research

The use of quantitative criteria is also motivated by the recognised limitations of intuition-based approaches. Both native speakers' and learners' intuitions have been shown to be unreliable for identifying collocations (Xiao, 2015), particularly medium- and low-frequency collocations (Siyanova-Chanturia & Spina, 2015). Corpus-based research therefore relies on quantitative evidence to characterise collocational patterns. As Gries (2022) notes, corpus-linguistic statistics are ultimately based on frequencies of occurrence and/or co-occurrence, which underpin three dimensions of statistical information: frequency, dispersion, and association.

Frequency provides information about how often a word combination occurs in a corpus and may be reported as tokens and/or types per se, and as raw or normalised frequency, depending on the purpose of the analysis. In collocational research, minimum frequency thresholds are commonly applied to ensure that sufficient corpus evidence is available for a word combination.

Dispersion concerns how occurrences of a linguistic feature are distributed across texts or other corpus segments. It has been described as a measure of “how (un)evenly this feature occurs across the corpus” and as “a proxy measure of occurrence regularity” (Gablasova et al., 2017, pp. 160-161). For instance, although *risk issues* (54 occurrences) and *moral issues* (51 occurrences) have similar raw frequencies in the BNC, they differ markedly in dispersion: the former expression occurs in a single text, whereas the latter appears across more than 41 texts (Gablasova et al., 2017). Therefore, collocations that occur across a wide range of linguistic settings are considered more general and more likely to be encountered by language users.

Association concerns the statistical relationship between co-occurring items. Association measures (AMs) are widely used to distinguish random co-occurrences from statistically associated word pairs. They assign an association score to each word pair, which can be used either to rank word pairs according to their strength of association or to select collocations by applying a predefined threshold (Evert, 2005). Many AMs derive this score by comparing “the observed co-occurrence frequency against the expected frequency under the null hypothesis of independence (i.e., complete absence of association)” (Evert, 2008, p. 1228). In contrast to approaches based on frequency, AMs “combine information about frequency with other collocational properties that can be expressed mathematically” (Gablasova et al., 2017, p. 159). Different AMs may therefore assign different scores to the same word pairs and, consequently, produce different rankings of collocational strength.

Table 1 presents the formulas for three AMs commonly discussed in collocation research: MI, t-score, and Log Dice. Of these, Log Dice is used in the present study. The **MI score** is widely used as “a measure of collocational strength and direction of association (i.e., attraction or repulsion between two lexical items)”. It is “computed by dividing the observed frequency of the co-occurring word in the defined span for the node word by the expected frequency of the co-occurring word in that span and then taking the logarithm of the result” (Xiao, 2015, p. 109). An MI of 3.0 or above has conventionally been used as a threshold for identifying significant collocations (Gablasova et al., 2017).

The **t-score** provides a different perspective by taking the amount of corpus evidence (i.e., corpus size) into consideration. It is calculated from “the difference between the observed and expected means, scaled by the variance, to determine the probability of a particular sample of that mean and variance under the assumption of the normal distribution of the dataset” (Xiao, 2015, p. 109). A t-score of 2.576 or above has conventionally been considered statistically significant.

Log Dice provides an alternative measure based on “the harmonic mean (a type of average appropriate for ratios) of two proportions that express the tendency of two words to co-occur relative to the frequency of these words in the corpus”. It operates on a scale with a fixed maximum value of 14 and is relatively stable across corpora of different sizes, making it useful

for cross-corpus comparison (Gablasova et al., 2017, pp. 164-165).

Although frequency-based criteria, dispersion measures, and AMs provide systematic quantitative evidence for characterizing collocations, their use also raises important methodological questions concerning how such measures are selected, interpreted, and applied. These issues are considered in the following subsection.

Table 1

Common Association Measures in Collocation Research

Association Measures	Formula
MI	$\log_2 \left(\frac{f_{AB}N}{f_A f_B} \right)$
T-score	$\frac{f_{AB} - \frac{f_A f_B}{N}}{\sqrt{f_{AB}}}$
Log Dice	$14 + \log_2 \frac{2 \cdot \ w_1, R, w_2\ }{\ w_1, R, *\ + \ \ast, \ast, w_2\ }$

Note. f_A denotes the frequency of the node word, f_B the frequency of the collocate, and f_{AB} the co-occurrence frequency of the node and the collocate. N represents corpus size. $\|w_1, R, w_2\|$ refers to the number of occurrences of the triple. $\|w_1, R, *\|$ refers to the number of occurrences of the first word in the grammatical relation with any second word. $\|\ast, \ast, w_2\|$ refers to the number of occurrences of the second word in any grammatical relation with any first word. Formulas adapted from *Statistics Used in Sketch Engine*, by Sketch Engine, n.d.

2.2.3.3 Methodological issues in collocation research

Two methodological issues are particularly relevant to the use of AMs in collocation research. The first concerns their selection and interpretation. Although widely used, AMs differ in their validity as measures of association, in their underlying assumptions, and in the aspects of the collocational relationship they capture. A limited understanding of these differences may have important consequences for the interpretation of collocational patterns in both L1 and L2 production (Gablasova et al., 2017). Accordingly, the choice of AM can influence which word combinations are identified as strongly associated and, consequently, the patterns that emerge from the analysis.

Given the large number of AMs available, researchers must decide which measure or combination of measures is appropriate for their analysis. Gries (2022) identifies validity as a central consideration in this decision and distinguishes two complementary aspects: whether an AM primarily captures the association it is intended to measure rather than other dimensions of statistical information, such as frequency, and whether its results correspond appropriately with relevant external evidence. The conflation of association with frequency becomes particularly problematic when AMs are used for linguistic, theoretical, or psycholinguistic interpretation rather than for purely descriptive purposes.

Gries (2022) further shows that commonly used AMs differ considerably in the extent to which they reflect association rather than co-occurrence frequency. MI reflects association relatively independently of frequency and therefore should not simply be characterised as a measure that favours infrequent combinations; frequent combinations can also receive high MI scores when their association is sufficiently strong. MI is thus relatively close to a genuine measure of association. By contrast, the t-score reflects predominantly frequency and relatively little association, making its interpretation as a measure of collocational strength more problematic. Log Dice performs comparatively well: while the unlogged Dice coefficient is more strongly

correlated with frequency than with association, the logged version is more strongly correlated with association than with frequency and is therefore closer to a genuine AM.

The interpretation of AMs is further complicated by their statistical assumptions and by the distributional characteristics of linguistic data. As Xiao (2015) notes, many statistical measures rely on assumptions of random distribution that do not adequately reflect the formulaic nature of language use. Linguistic frequency distributions are also highly skewed, as Biber et al. (2021) observe, with a relatively small number of items accounting for a large proportion of occurrences. Such distribution asymmetry affects both the occurrence of particular collocations in language and speakers' exposure to and experience with them.

Frequency, however, represents only one aspect of collocational behaviour. Although both L1 and L2 speakers are sensitive to the frequency distribution of word combinations, frequency alone does not fully account for collocational knowledge, which may also be shaped by other cognitive predictors, including "proficiency of learners and salience, uniqueness, semantic meaning, or personal relevance of expressions" (Gablasova et al., 2017, pp. 159-160). This limitation of frequency as a sole indicator motivates consideration of other distributional properties of word combinations in corpus-based analysis, including dispersion, *exclusivity*, and *directionality*.

According to Gablasova et al. (2017), **exclusivity** refers to "the extent to which the two words appear solely or predominantly in each other's company" and can be expressed as the relationship between "the number of times when they are seen together as opposed to the number of times when they are seen separately in the corpus" (p. 160). High exclusivity has also been correlated positively with salience and noticing. Examples include *zig zag* and *okey dokey*, whose component words occur highly exclusively with one another. **Directionality**, by contrast, postulates that the strength of association is not necessarily equal in both directions. In other words, "each of the two words in the collocation involves different degree of probability that it will occur with the other word in the pair" (pp. 160-161). For example, in the BNC, 11% of occurrences of *decorations* are preceded by *Christmas*, whereas only 0.5% of occurrences of *Christmas* are followed by *decorations*.

These distributional properties can be captured to varying degrees by different quantitative measures. For example, t-score emphasizes adjusted frequency, MI highlights rare exclusivity, whereas Log Dice captures exclusivity more generally. In addition, directionality can be captured by Delta P, whereas dispersion can be grasped by Cohen's d (Gablasova et al., 2017). The variation underscores the need to select and interpret measures according to the particular aspect of collocational behavior relevant to the research question.

A second issue concerns the extent to which association scores are influenced by the corpus from which they are calculated. Association scores may vary across corpora because the frequencies on which they are based are affected by the characteristics of corpora, including size, representativeness, register, and mode (Gablasova et al., 2017). This issue is particularly relevant in L2 research, where association scores derived from an L1 reference corpus are often used to characterize word combinations occurring in learner data.

In learner corpus research, each word combination extracted from learner data is typically searched for in a native reference corpus, such as the BNC or COCA. Combinations that meet a predefined frequency threshold in the reference corpus are assigned association scores (e.g., MI and t-score), which are then used as indicators of collocational strength; combinations falling below the threshold are classified as below threshold. The resulting association scores can then be analysed in different ways. One approach is to group them into predefined bands or categories. For example, Durrant and Schmitt (2009) proposed seven bands for both MI and t-score, which were subsequently grouped into broader *high* and *low* categories, while Granger and Bestgen (2014) opted for a four-way classification (i.e., non-collocational, low, medium and high). In

both studies, the proportions of word combinations falling into the respective categories were compared across proficiency groups using independent-samples t-tests to determine whether the distributions differed significantly across proficiency groups. Such categorical approaches have, however, been criticised for reducing continuous association scores to discrete groups.

An alternative approach is to retain association scores as continuous values and summarize them at the text level using measures of central tendency. Bestgen and Granger (2014), for example, used CollGram, a technique that assigns two association scores to each bigram (i.e., “any contiguous pair of words in the L2 texts”) in a learner text on the basis of a large reference corpus (p. 31). They then calculated mean MI and mean t-scores, together with the proportion of bigrams absent from the reference corpus, to construct CollGram profiles of learner texts and relate them to the development of L2 writing proficiency. Such approaches enable systematic comparisons between native and non-native production, thereby offering insight into learners’ use of phraseological units. Empirical findings consistently indicate that L2 learners tend to overuse highly frequent collocations while underusing less frequent but strongly associated items. As the latter are highly salient to native speakers, their relative absence in learner production may contribute to the perception that non-native writing is devoid of *idiomaticity* (Durrant & Schmitt, 2009).

However, much early research relied on large general-purpose reference corpora such as the BNC and COCA because of their size, balance, and representativeness. As these corpora comprise a wide range of spoken and written registers, they may not always be directly comparable with learner corpora representing a more specific register. Variation also occurs within the same register. For example, Lee (2025) shows that collocational use in L1 academic writing varies across writers with different ages and levels of experience in academic English. Variation both across registers and among texts within the same register highlights the importance of comparability between the reference corpus and the learner corpus under investigation.

One way of improving such comparability is to calculate association scores from a more specialised reference corpus that more closely matches the learner corpus. Paquot (2019), for example, used the L2 Research Corpus (L2RC), a corpus of expert academic writing, to calculate MI scores for word combinations occurring in learner research papers in linguistics. In this case, the reference corpus was more closely in line with the learner corpus in terms of register, providing a more comparable basis for evaluating L2 phraseological use. This illustrates why the characteristics of the reference corpus need to be considered when association scores are used to compare L1 and L2 phraseological patterns. The same consideration also applies to the operationalisation of phraseological sophistication. As extrinsic and genre-sensitive measures, indices of phraseological sophistication can capture learners’ command of context-specific phraseology only insofar as the collocation lists and reference corpora on which they are based are appropriate to the target context. This is discussed further in Section 2.2.4.2.

2.2.4 The role of collocations in language use and research

The importance of collocations extends well beyond the simple repeated co-occurrence of words in texts: collocations contribute to meaning construction in discourse, shape language processing and production, and serve as indicators of language proficiency. This centrality has made collocations a major focus across multiple strands of linguistic research over the past decades (Gablasova et al., 2017; Szudarski, 2023). In particular, advances in corpus linguistics have enabled the systematic identification and quantitative analysis of co-occurrence patterns that were largely inaccessible to earlier introspection-based approaches (Xiao, 2015), while also supporting the development of collocational resources and analytical methods. This section considers the significance of collocations from three perspectives: meaning construction, L2

learner performance, and corpus-based resources and analytical approaches.

2.2.4.1 The role of collocations in meaning construction

From the perspective of meaning construction, the role of collocations in discourse can be captured through two distinct yet interdependent concepts of semantic preference and semantic prosody. Semantic preference refers to the tendency of a lexical item to co-occur with a set of semantically related words, while semantic prosody arises from the interaction between a node and its typical collocates, and primarily functions to express attitude or evaluation (Louw, 2000; Partington, 2004; Stubbs, 2002). Partington (2004) provides an illustrative example with the verb *undergo*, which shows a semantic preference for lexical sets of “change”, “medicine”, “testing” and “involuntariness”, thereby contributing to an overall unfavourable semantic prosody (pp. 150-151).

Partington (2004) further notes that the two concepts differ in their scope: semantic preference can be treated as a feature of the collocates, whereas semantic prosody is more appropriately viewed as a feature of the node word. In this sense, semantic preference “contributes powerfully to building” semantic prosody, and semantic prosody, in turn, “dictates the general environment which constrains the preferential choices of the node item” (Partington, 2004, p. 151). More broadly, the relationship between lexical units can be described at four levels from the lowest to the highest: collocation (“the relationship between a node and individual words”), *colligation* (“the relationship between a node and grammatical categories”), semantic preference (“semantic sets of collocates”), and semantic prosody (“affective meaning of a given node with its typical collocates”) (Sinclair, 1996, 1998; Stubbs, 2001b, as cited in Xiao, 2015, p. 114).

Building on this, collocations are integral to language comprehension and production. By constraining the co-occurrence of lexical items, they facilitate the disambiguation of polysemous words and, together with semantic prosody, guide language users in selecting contextually appropriate *near-synonyms* (Xiao, 2015). For instance, Phoocharoensil (2025), using data from COCA, shows that near-synonyms such as *poisonous* and *venomous* exhibit systematic differences in their collocational patterns. Both adjectives are more prevalent in relatively informal contexts (i.e., fiction and magazines), but differ in their typical noun collocates, with *poisonous* tending to co-occur with nouns related to gases, chemicals, substances, and food, and *venomous* more often collocating with nouns denoting actions, emotions, and food items.

Overall, collocations provide an essential link between lexis and discourse, demonstrating that meaning is not only encoded in individual words but also emerges from their co-occurrence.

2.2.4.2 The role of collocations in L2 learner performance

Beyond their role in meaning construction, collocations are widely recognized as a key dimension of L2 proficiency. In learner corpus research and studies of linguistic complexity, variation in collocational patterns provides important insights into learners’ language use and development.

Learner corpus research

Empirical findings consistently show that the development of L2 phraseological competence is gradual and relatively slow. Collocational errors persist across proficiency levels, and learners’ production is typically characterized by underuse, overuse, and various forms of misuse of collocations in comparison with native-speaker norms (Paquot & Granger, 2012). More specifically, learners display uneven patterns in collocation density (i.e., the number of collocation tokens) depending on syntactic categories, such as the overuse of verb + noun collocations and underuse of adjective + noun collocations (Vincze et al., 2016, as cited in Granger, 2018). Regarding collocation diversity (i.e., the number of collocation types), learners’ production tends to be limited, with a strong reliance on high-frequency collocations (e.g., *make + decision*) or

rather fixed combinations (e.g., *take + part in*) (Wang, 2016, as cited in Granger, 2018). With respect to accuracy and appropriacy, one convergent result is that the proportion of deviant collocations is relatively high. Such difficulties are influenced by factors including L1 transfer, typological differences, and the distance between the learners' L1 and L2 (Granger, 2018).

Longitudinal research further suggests that collocational development progresses slowly, with limited gains observed over short instructional periods (e.g., one semester or one year). This limited development is also reflected in quantitative studies using association measures. For example, in Bestgen and Granger's (2014) study of bigrams over one semester, no significant increase was observed in the MI scores of verb-noun statistical collocations, while a decrease was found in the number of collocations identified by t-scores. This pattern may be related to the frequency of collocations in learners' input, as the acquisition of collocational knowledge depends on repeated exposure to frequent co-occurrence patterns (Granger, 2018).

Given these persistent difficulties and the slow pace of development, a strong command of collocations is widely regarded as a key indicator of native-like proficiency and fluency, as it enables speakers and writers to process and produce language more efficiently while conveying complex meanings (Xiao, 2015). Accordingly, there has been increasing interest in capturing collocational competence within the broader construct of linguistic complexity, which aims to "investigate L2 performance, track its development, and to assess its proficiency" (Spina, 2025, p. 37).

Phraseological complexity research

Within this line of research, the construct of phraseological complexity has been proposed to systematically theorize and operationalise linguistic complexity at the level of word combinations (Paquot, 2019; Paquot & Naets, 2025).

The construct of phraseological complexity proposed by Paquot (2019) is conceptualised along two dimensions: *diversity* and *sophistication*. It is defined as "the range of phraseological units that surface in language production and the degree of sophistication of such phraseological units" (p. 124). The latter dimension—phraseological sophistication—has been further re-conceptualized as a multi-dimensional construct comprising association strength, register specificity, and frequency (Paquot & Naets, 2025). Accordingly, "a learner text with a wider diversity of (target-like) word combinations and a higher proportion of sophisticated units is considered to be more complex than one where the same set of (more frequent) word combinations is often repeated" (Paquot & Naets, 2025, p. 218).

Empirically, phraseological units are approached via co-occurrences in three dependency relations, namely adjectival modifiers (adjective + noun), adverbial modifiers (adverb + adjective, adverb or verb), and verb + direct object. Diversity is operationalised through "a (derived) type-token ratio representing the number of unique phraseological units to the total number of phraseological units" (Paquot, 2019, p. 125). In contrast, the association dimension of sophistication is operationalized through "pointwise mutual information (MI) scores associated to word combinations" (Paquot, 2019, p. 126). The register specificity dimension of sophistication is measured as "the ratio of academic collocations to the total number of collocations for each type of relational dependencies" (Paquot & Naets, 2025, p. 227), while the frequency dimension of sophistication is measured as the proportion of dependencies in a learner text that fall outside the most frequent combinations (approximately the top 80% coverage) in a reference corpus, with thresholds adjusted for each dependency type (Paquot & Naets, 2025).

This framework is of great significance because it provides empirically grounded phraseological measures for describing L2 performance, assessing proficiency, and tracing development across modes and tasks in a range of L2 contexts. Notably, measures of phraseological sophis-

tication have been shown to outperform traditional indices of syntactic and lexical complexity in distinguishing advanced proficiency levels, highlighting their value for language assessment (Paquot, 2019; Paquot & Naets, 2025; Spina, 2025).

More specifically, association-based measures of sophistication—particularly those based on verb + direct object dependencies—display a significant linear increase across proficiency levels (i.e., from B₂ to C₁ to C₂), highlighting their robustness as indicators of advanced proficiency. Similarly, measures of sophistication by register specificity (academic verb + direct object combinations) not only differentiate higher-level learners, but also correlate with association-based measures. These findings underscore the significance of the verb + direct object relation as a lexico-grammatical context that remains challenging for learners in academic discourse, even for advanced learners (Paquot & Naets, 2025).

Nevertheless, variation in learners' phraseological performance has been shown to depend on contextual factors, including task conditions and the syntactic types of the phraseological units examined (Spina, 2025). Moreover, Paquot and Naets (2025) emphasise the need to explore “the distribution of co-occurrences across a range of reference corpora” in order to develop appropriate context-sensitive resources that could be used to examine “phraseological sophistication in relation to frequency” (p. 239). This is particularly important when learner data extend beyond academic contexts, where existing academic collocation lists may not adequately capture register-specific usage, thus necessitating collocation datasets derived from more specialized reference corpora (Paquot & Naets, 2025).

Collectively, these findings position collocations as both sensitive indicators of L2 development and context-dependent patterns whose distribution varies across communicative settings. This draws attention to the need to account for the distribution and variation in collocational patterns across different contexts, particularly when evaluating phraseological competence and complexity. Failure to consider such variation may undermine the practice of treating L1 users' usage as a stable norm or reference point against which L2 learners' collocational patterns are evaluated (Lee, 2025).

2.2.4.3 Collocational resources and analytical approaches

Collocational information has been used in corpus research both to develop lexical and phraseological resources and to support the analysis of patterns of language use. Two broad strands are particularly relevant here: the development of collocational resources for language learning and lexicography, and corpus-based approaches to analysing and representing collocational patterns.

Collocational resources for learning and lexicography

One prominent strand of research concerns the development of word and collocation lists designed to identify items of pedagogical relevance in particular contexts. These resources range from general academic lists (e.g., Ackermann and Chen's (2013) Academic Collocation List) to discipline-specific resources targeting particular academic domains (e.g., Gholaminejad and Sarab's (2020) word and collocation list for language teaching and applied linguistics). While such studies have substantially advanced the description of phraseology in academic discourse, they also reveal a strong bias towards academic registers, with comparatively less attention to professional and non-academic contexts (Conrad, 2015).

Beyond the identification of individual words and collocations, research has increasingly focused on lexico-grammatical patterns. Hunston and Francis (2000) define the *pattern* of a word as “all the words and structures which are regularly associated with the word and which contribute to its meaning” (p. 37). A pattern is identified when a combination occurs with sufficient frequency, is associated with a particular word choice, and carries a distinct meaning. This perspective extends lexical description from knowledge of individual words to knowledge

of the grammatical and phraseological environments in which they typically occur. From a pedagogical perspective, it therefore relates closely to the distinction between vocabulary breadth (the number of words known) and vocabulary depth (the extent of knowledge about each word) (Schmitt et al., 2011). Knowledge of such lexico-grammatical patterns constitutes an important aspect of vocabulary depth.

This perspective has been operationalised in corpus-based resources. Ma and Qian (2020), for example, developed a grammar pattern list for the 115 most frequent academic verbs and provided frequency information for each identified pattern. For the V + N pattern, for instance, 100 verb types were identified, with frequencies ranging from 7 for *tend* to 1,339 for *contain*. The study therefore highlights the pedagogical value of incorporating information about frequent lexico-grammatical patterns into vocabulary instruction in English for Academic Purposes (EAP). In a similar vein, Green (2019) used collexeme analysis (Stefanowitsch & Gries, 2003)—a method for identifying lexical items that are strongly associated with a particular slot in a single grammatical construction. The findings show that incorporating lexico-grammar patterns and statistical information (i.e., pattern frequency and collexeme association strength) can enrich traditional wordlists and inform more effective vocabulary teaching in EAP.

These developments are also reflected in corpus-based dictionaries that provide learners with detailed phraseological information. One example is the Louvain English for Academic Purposes Dictionary (LEAD), a web-based EAP dictionary designed by Sylviane Granger and Magali Paquot for non-native writers (Granger & Paquot, 2010, 2015; Paquot, 2012). It provides corpus-based descriptions of non-technical academic words and phrases, with particular attention to their collocations and recurrent phrases. Users can explore the phraseological environment of a search word through authentic example sentences extracted from specialized corpora.

Alongside learner-oriented dictionaries, corpus-driven lexicography focuses on the systematic description of lexical patterns and meanings. One prominent example is the *Pattern Dictionary of English Verbs* (PDEV) (<http://pdev.org.uk>) (Hanks, 2013), which is the principal output of Corpus Pattern Analysis (CPA) (Hanks, 2004). Drawing primarily on the written component of the BNC (BNC50), PDEV describes the typical patterns of over 1,600 verbs (Baisa et al., 2015), presents these patterns using syntactic and semantic categories, and links them to examples from corpus data.

These resources demonstrate how corpus-based collocational knowledge can be organised for pedagogical and lexicographic purposes. Despite these advances, existing descriptions of lexical-grammatical patterns often focus on individual registers and are rarely situated within a systematic comparative framework. This limitation is also acknowledged by Green (2019) and Ma and Qian (2020), who call for broader empirical coverage. As a result, it remains unclear how such patterns—especially verb-collocate combinations—vary across communicative contexts and which patterns are register-specific or more widely shared.

Corpus-based analysis and representation of collocational patterns

A related line of research uses collocational information to examine lexical and phraseological patterning directly in corpus data. CPA, the methodology underlying PDEV, provides one such framework for analysing the relationship between recurrent patterns and meaning in corpus data. As Baisa et al. (2015) explain, CPA “associates word meaning with word use by means of analysis of phraseological patterns and collocations” (p. 75). Within this framework, verb meanings are identified through “a combination of valency and collocates in which such a combination has a distinctive meaning” (Hanks, 2012, p. 58). Starting from the assumption that a verb determines how many other elements (or complements) are required to form a grammatical sentence, Herbst (2004) defines valency as “the number of complements a verb takes”, with these complements

dependent upon the governing verb (p. xxiv). CPA further distinguishes among semantic types, contextual roles, and lexical sets to capture fine-grained distinctions in meaning and usage.

Studies applying CPA show that the relationship between verb patterns and meanings can vary across disciplinary contexts and between L1 and L2 usage. Rees (2018), for instance, analysed 30 high-frequency verbs in research articles from history, microbiology and management studies, and found that the prototypical meanings of individual verbs varied across disciplinary domains in relation to their syntagmatic contexts. Alqarni (2019), by contrast, examined sixteen verbs in two sections of the International Corpus of Learner English and compared their use with native-speaker data in the Louvain Corpus of Native English Essays (LOCNESS). The analysis identified associations between verb patterns and meanings in both L1 and L2 data, as well as patterns that were divergent, underrepresented, or overrepresented in learner production.

A different means of analysing collocational relationships is provided by collocation graphs and networks. A collocation graph “shows the relationship between the node and its collocates as measured by a particular collocation statistic (association measure). The length of the link (edge) between the node and the collocates is inversely proportional to the strength of association: the stronger the association, the shorter the link” (Brezina, 2018, pp. 62-63). Such a simple collocation graph can be expanded into a more complex collocation network which is “the best form of display of connected collocations” because it includes “multiple nodes, their respective unique collocates, as well as shared links and shared collocates” (Brezina, 2018, p. 60).

Compared with traditional tabular representations, collocation graphs and networks can not only “indicate complex relationships between multiple nodes”, but also “show meaning associations, cross-associations and shared as well as unique collocates” (Brezina, 2018, p. 61). Brezina (2018), for example, used collocation graphs to investigate the development of phraseological complexity of combinations that include *make*, *take* and *do* in the Trinity Lancaster Corpus (TLC) of spoken L2 English. Higher proficiency was associated not only with a larger number of collocates, but also with a shift towards more semantically integrated (e.g., *make + sure*) and more abstract combinations (e.g., *take + advantage*) (p. 73). Such approaches highlight the potential of collocation networks to uncover patterns of linguistic complexity and development.

Across collocational resource development and corpus-based analytical studies, a common limitation is the tendency to focus on single disciplines or registers, with limited systematic comparison across communicative contexts. This gap points to the need for an approach that can account systematically for contextual variation in language use. In this respect, the register perspective proposed by Biber and Conrad (2019) is well suited to this purpose because it relates the description of the situational characteristics to the functional interpretation of why particular linguistic features tend to occur in those contexts. Accordingly, the study of register variation focuses on the functional associations between linguistic patterns and situational factors (Biber & Conrad, 2019). The following section therefore introduces the register perspective and major approaches to register variation, providing a basis for analysing collocational patterns across registers.

2.3 Register

2.3.1 Register and its analytical framework

Research on variation in language use has a long tradition in linguistics (Seoane & Biber, 2021). Given the reliance of empirical register analysis on collections of texts, the development of register research has been closely intertwined with advances in corpus linguistics. Although early corpus-linguistic studies did not fully exploit the potential of large-scale corpora for

describing registers and register variation, systematic corpus-based investigations of variation across registers began to emerge in the 1990s (Seoane & Biber, 2021). Since then, distinguishing text types or registers through computational and statistical methods has become a central pursuit in corpus-linguistic research (Gries et al., 2011; Spina & Tanganelli, 2012).

Conceptualising register

Within this line of research, the concept of *register* has been developed to refer to “varieties associated with particular situational contexts that can be characterized for their audiences, mediums (e.g., spoken or written mode), interactivity, production circumstances, communicative purposes” (Seoane & Biber, 2021, p. 15). This definition highlights the central role of context in shaping language use, as speakers and writers adapt their linguistic choices to the communicative demands of specific situations. Varieties associated with particular situations of use and particular communicative purposes are often referred to as either *registers* or *genres*. Biber and Conrad (2019) distinguish these as two perspectives on the analysis of texts. The genre perspective focuses on linguistic characteristics that are used to structure complete texts, whereas the register perspective focuses on pervasive features that occur throughout texts (see Table 2). This distinction is particularly relevant for the present study, which focuses on lexico-grammatical features rather than “specialized expressions, rhetorical organization, and formatting” (p. 16). Accordingly, the term *register* is adopted throughout this study to refer to varieties associated with particular situations of use and communicative purposes.

Table 2

Defining Characteristics of Registers and Genres, adapted from Biber & Conrad (2019, p. 16)

Defining characteristic	Register	Genre
Textual focus	Sample of text excerpts	Complete texts
Linguistic characteristics	Any lexico-grammatical feature	Specialized expressions, rhetorical organization, formatting
Distribution of linguistic characteristics	Frequent and pervasive in texts from the variety	Usually once-occurring in the text, in a particular place in the text
Interpretation	Features serve important communicative functions in the register	Features are conventionally associated with the genre: the expected format, but often not functional

More broadly, register variation is widely recognized as a universal feature of language use, as all cultures use languages in different situations for different communicative purposes. Given the ubiquity of register variation, understanding how linguistic features are distributed in patterned ways across registers is central to both the description of language and the development of cross-linguistic theories of language use (Biber & Conrad, 2019).

Analytical framework and procedures for register analysis

On a theoretical level, a comprehensive description of a register involves three components: “(1) the situational/communicative description, (2) the description of pervasive linguistic features, and (3) the analysis of the functional associations between linguistic forms and situational contexts” (Biber & Conrad, 2019, p. 8). Among these components, situational characteristics are considered more basic because they are not derived from linguistic phenomena, whereas linguistic differences across registers can be explained in relation to differences in situational context.

The linguistic features associated with a register are both pervasive and functional. They

are pervasive in that they occur relatively frequently throughout the register, reflecting their suitability to the communicative purposes and situational contexts of that register. Meanwhile, they are functional in that their distribution varies systematically across contexts: they are used to greater or lesser extents depending on the situational context, rather than occurring randomly. As a result, systematic patterns of register variation arise from the functional relationship between linguistic forms and their situations of use. Functional interpretation, in particular, is central to register variation analysis, as it seeks to explain observed linguistic differences in relation to differences in situational characteristics (Biber & Conrad, 2019).

On a practical level, register analysis proceeds in three main stages. The first involves identifying and describing situational characteristics, including “participants, relations among participants, channel/mode, processing circumstances, setting, communicative purposes and topic”, as outlined in the framework proposed by Biber and Conrad (2019, p. 40). The second step concerns the description of the typical (pervasive) linguistic features that may distinguish registers. To guide the selection of relevant features, Biber and Conrad (2019) proposed a set of 15 broad categories of linguistic characteristics that have proven useful in previous register studies. These may include verb features such as verb valencies, semantic classes of verbs, and action verbs with inanimate subjects; features of noun phrases, such as semantic category of noun; features of adverbials, such as circumstance and stance adverbial types; and general vague words, such as *thing*, *stuff*, which are particularly characteristic of conversation (pp. 65-68).

The third step focuses on functional interpretation, which moves beyond description to explain why particular linguistic features are associated with particular situational characteristics and why they are especially common in those contexts. By relating the situational and linguistic descriptions, functional explanations can be formulated for the observed linguistic differences. To support this stage, Biber and Conrad (2019) summarize 15 major functions that distinguish among registers, together with selected linguistic features that have been found to be associated with these functions.

These functions include **interactivity**; **personal stance**, associated with features such as possibility adverbs; **referring to the time and place of communication**, associated with expressions such as *here*, *there*, *yesterday*, and *last week*; **referring to shared personal knowledge**, associated with first names of friends and vague references such as *thing*; and **referring to shared expert/professional knowledge**, associated with technical words and phrases and references to published research studies.

The framework also identifies five general communicative purposes: **narrative**, associated with features such as place and time adverbs; **description**, associated with adjectives, adverbs, and stative verbs such as *be*, *seem*, *appear*, *look*; **directive**, associated with features such as desire verbs with *you*; **procedural**, associated with ordinal numbers such as *first*, *second*; **explanatory/expository**, associated with nouns and attributive adjectives. In addition, three functions concern the presentation of information—**elaborating information**, **condensing information**, and **marking logical relations**—while two functions concern production circumstances: **real time** (associated with vague nouns such as *thing*) and **careful production and revision** (p. 73).

Application to selected registers

To illustrate the application of their analytical framework, Biber and Conrad (2019) provide detailed descriptions of the situational characteristics of news reports and academic prose (pp. 113-115). Building on these descriptions, the present study summarises the situational characteristics of four registers—conversation, parliamentary debates, news reports, and academic prose—with the first two added on the basis of the same analytical framework, as shown in Table 3.

These registers differ systematically along key situational dimensions, including communicative purpose, degree of interactivity, and mode of production. Conversation is typically interactive and dialogic, whereas news reports and academic writing are predominantly expository and monologic. Parliamentary debates occupy an intermediate position: they are primarily argumentative and partly dialogic, while also exhibiting features of planned discourse. Although delivered in spoken form (or more accurately, “written-to-be-seen-as-if-not-written” rather than spontaneous), they often rely on previously written scripts and thus combine features of both spoken interaction and written discourse, resulting in a hybrid mode (Dupont & Zufferey, 2017, p. 276). Table 3 thus provides a basis for comparing these registers in terms of their situational characteristics.

In addition to situational analysis, Biber and Conrad (2019) also compare the distribution of 15 selected linguistic features across news, academic prose and conversation (pp. 118-119). Drawing on their findings, Table 4 provides a reduced set of features directly relevant to the present study, showing that verb-related features vary systematically across registers: conversation is characterised by frequent present tense forms and modals, whereas news exhibits a stronger preference for past tense, and academic prose makes extensive use of passive constructions (about 25% of all finite verbs). Similar variation can be observed in other linguistic features that frequently interact with verbs. Nouns are common in news and academic prose, while personal pronouns are particularly frequent in conversation. Time adverbials are especially prominent in news, whereas both time and place adverbials occur frequently in conversation but are comparatively rare in academic prose.

These distributional patterns can be interpreted in functional terms. For instance, the prevalence of past-tense verbs and time adverbials in news is consistent with its *narrative* function, which typically involves reporting events in time. In contrast, the frequent use of passive constructions and nouns in academic prose reflects an orientation towards *explanatory* or *expository* functions, where information is presented in a more abstract and impersonal manner. In conversation, the high frequency of present tense forms, modals, and personal pronouns (+ mental or desire verbs) serves functions such as *interactivity* and the expression of *personal stance*, as evidenced by expressions like *I think* (p. 73).

Together, the situational characteristics of the four registers and the distributional patterns of selected linguistic features provide an empirical basis for the present study. They establish key contextual differences among the registers and illustrate how these differences are reflected in language use, thereby providing a foundation for analysing and interpreting variation in verb collocations across registers. To examine such variation systematically, it is also necessary to consider how register variation is conceptualised and operationalised in previous research.

Table 3*Key Situational Differences among Four Registers*

Situational characteristic	News reports	Academic prose	Conversation	Parliamentary debates
I. Participants				
A. Addressor				
1. Single / plural / institutional / unidentified	may be single, plural, institutional or unidentified	usually singular or plural, sometimes institutional	usually single (per turn); identifiable individuals	single or plural; identifiable but institutionally defined (Members of Parliament, ministers)
2. Social characteristics	often adult journalist but varies with sub-register	often adult trained professional but varies with sub-register (e.g., can be student)	varies (e.g., friends, family)	adult professionals with institutional roles; varies by political position
B. Addressee				
1. Single / plural / unenumerated	group (very general)	group (more specialized than for news reports, including other professionals in the academic field, students, etc.)	single or small group; identifiable participants	plural identifiable participants (other MPs) plus broader unenumerated audience (public)
2. Self / other	other	other	other	other
C. Onlookers?	N/A	N/A	usually none	present (e.g., public, media observers)

Continued on next page

Situational characteristic	News reports	Academic prose	Conversation	Parliamentary debates
II. Relations among participants				
A. Interactiveness	no direct interaction, although readers can post comments on many news websites	no direct interaction	direct interaction; rapid turn-taking	direct interaction; turn-taking regulated by institutional procedures
B. Social roles	varies	varies	relatively equal; roles negotiated locally	institutionally defined; hierarchical or role-based
C. Personal relationship	none	none	typically close (e.g., friends)	none or distant; professional and often adversarial
D. Shared knowledge	varies	varies	high shared personal knowledge; situation-dependent	shared institutional/political knowledge; less personal shared context
III. Channel				
A. Mode	writing – can be printed or in electronic format available online	writing	speech	speech (often transcribed)

Continued on next page

Situational characteristic	News reports	Academic prose	Conversation	Parliamentary debates
B. Specific medium	printed and/or online	printed and/or online	face-to-face, telephone	face-to-face in institutional setting; recorded and archived
IV. Processing circumstances				
A. Production	time for planning, revising, editing (often includes tight deadlines); often have strict space constraints	time for planning, revising, editing; space constraints vary	real-time production; little opportunity for planning or revision	real-time production; may include prepared or scripted contributions
B. Comprehension	varies depending on reader – careful reading or skimming; opportunity for re-reading	often careful reading but may be skimmed quickly; opportunity for re-reading	real-time processing; no opportunity for re-reading	real-time processing for participants; transcripts allow later re-reading
V. Setting				
A. Time and place shared by participants?	no physically shared time or place but expected to be read on same day as produced	no shared time or place	yes (co-present participants)	yes (shared institutional setting)
B. Place of communication				

Continued on next page

Situational characteristic	News reports	Academic prose	Conversation	Parliamentary debates
1. Private / public	public (available for others to view)	public (available for others to view)	usually private or semi-private	public (institutional setting)
2. Specific setting	usually associated with a specific city; may be read more widely; certain articles may be produced by a wire service with no specific setting	no specific setting	varies (e.g., home)	fixed institutional setting (parliament chamber)
C. Time period	contemporary (in this study)	contemporary (in this study)	contemporary (in this study)	contemporary (in this study)
VI. Communicative purposes				
A. General purposes	informational – reports events; some articles do analysis; editorials and reader comments are likely to be much more opinionated	informational – inform and explain/interpret	interpersonal – maintain relationships; express attitudes and feelings	persuasive/argumentative – debate policies, justify positions

Continued on next page

Situational characteristic	News reports	Academic prose	Conversation	Parliamentary debates
B. Specific purposes	varies, with informing, reporting/narrating, and personal opinion purposes (e.g., news reports report daily events, editorials and reader comments express opinions and seek to persuade)	varies, with informative, explanatory, and persuasive purposes (e.g., research articles present new findings, but most academic texts also try to persuade readers about an interpretation of empirical findings)	varies (e.g., sharing experiences)	varies (e.g., arguing for/against policies, criticising opponents, influencing public opinion)
C. Purported factuality	factual reports separated from opinion pieces	factual with interpretation	mixed (personal experience, opinion, some factual reference)	mixed (factual claims combined with evaluative argumentation)
D. Expression of stance	varies – not expected to be overt in news reports, but common in editorials and reader comments	varies by discipline	frequent; overt expression of personal stance	frequent; often explicit and rhetorically marked (evaluation)
VII. Topic				
A. General topical domain	current newsworthy events in many areas; varies with different sections of paper	varies	everyday life, personal concerns	politics, governance, public policy
B. Specific topic	varies	varies	varies; often shifts rapidly and locally	more constrained; determined by agenda or debate topic

Table 4

Distribution of Selected Linguistic Features Relevant to Verb Collocations across Registers
(adapted from Biber & Conrad, 2019, pp. 118–119)

Linguistic feature	News	Academic prose	Conversation
Verb characteristics			
Present tense	less common than in academic prose; slightly more common than past tense	more common than in news; far more common than past tense	very common
Past tense	much more frequent than in academic prose; slightly more common than in conversation	rare	uncommon
Modals	uncommon; slightly less common than in academic prose; <i>will</i> and <i>would</i> most common	uncommon; slightly more common than in news; <i>can</i> and <i>may</i> most common	more common than in news or academic prose (about 15% of all finite verb phrases)
Passives	about 15% of all finite verbs	more common than in news; about 25% of all finite verbs	rare
Nominal features			
Nouns	very common; even more common than in academic prose	very common	less common
Personal pronouns	slightly more common than in academic prose, but still uncommon	rare	extremely common
Circumstance adverbials of time and place			
Time adverbials	by far most common	rare	common
Place adverbials	also common	rare	common

2.3.2 Conceptualising and analysing register variation

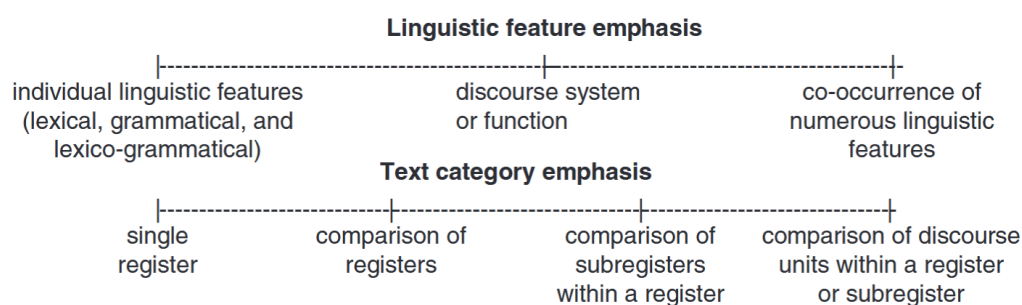
According to Seoane and Biber (2021), approaches to register studies can be broadly divided into two orientations. The first treats register as a predictor of linguistic variation, using it to account for variation in linguistic choices (e.g., the alternation between the *of*-genitive and the *'s*-genitive). From this perspective, linguistic features are treated as dependent variables whose distribution is conditioned by register. This orientation is characteristic of variationist research, which models probabilistic patterns of linguistic variation in relation to contextual factors. The

second orientation takes register itself as the primary object of description. This perspective underpins research traditions such as Language for Specific Purposes (LSP) and English for Specific Purposes (ESP), which describe the lexico-grammatical characteristics of particular registers in specific professional or disciplinary contexts.

Within the second orientation, studies may focus on either the description of individual registers or on the comparison of linguistic patterns across registers. Studies of register variation likewise describe the lexico-grammatical characteristics of texts, but place greater emphasis on comparing registers and interpreting the patterns of similarities and differences that emerge from them. Across such studies, there is a recognition that register variation is fundamental to understanding language use, and differences across registers are reflected not only in the frequency of pervasive linguistic features but also in their functional importance within specific communicative contexts (Seoane & Biber, 2021).

Methodologically, corpus-based studies of register variation can be situated within a framework proposed by Conrad (2015), which conceptualises research along two intersecting continua: linguistic feature emphasis and text category emphasis (see Figure 1). The first linguistic feature continuum concerns the level of linguistic analysis. At one end, studies focus on individual linguistic features. At the intermediate points, attention shifts to broader discourse systems or functions, which examine how specific systems or functions are realised across registers, such as the expression of future time, grammatical complexity, and expressions of stance or evaluation. At the far end, research examines the co-occurrence of numerous features, most notably through multidimensional (MD) analysis. The text category continuum concerns the scope of comparison, ranging from a single register to comparisons across registers, sub-registers, or discourse units within a register or sub-register. Studies can therefore be positioned at different points along both continua according to their analytical focus (p. 311).

Figure 1: *Linguistic Feature and Text Category Emphases in Register Variation Studies*



2.3.2.1 Empirical findings of studies using the individual feature approach

At the individual-feature end of Conrad's (2015) framework, register variation is examined through the distribution of particular lexical, grammatical, and lexico-grammatical features across text categories. A prominent and comprehensive example is the *Longman Grammar of Spoken and Written English* (Biber et al., 1999), which reports over 300 investigations of linguistic features across conversation, fiction, newspapers, and academic prose. These analyses range from simple frequency comparisons (e.g., the distribution of lexical word classes across registers) to more detailed grammatical and lexico-grammatical studies, while also considering discourse factors associated with linguistic choices. Notably, among the individual linguistic features investigated, research specifically examining verb use across registers remains relatively limited, with much of the detailed evidence provided by Biber and his colleagues. The discussion below therefore summarises findings most relevant to the present study, particularly variation in

verb-related features across registers.

Variation in high-frequency verbs and common lexical verbs

In the revised *Grammar of Spoken and Written English* (Biber et al., 2021), lexical verbs are classified into seven major semantic domains: “*activity verbs, communication verbs, mental verbs, causative verbs, verbs of simple occurrence, verbs of existence or relationship, and aspectual verbs*” (p. 359). The authors identify the twelve most common lexical verbs, each occurring over 1,000 times per million words. These verbs fall into three main semantic domains: six activity verbs (*get, go, make, come, take, give*), five mental verbs (*know, think, see, want, mean*) and one communication verb (*say*) (p. 371). These verbs occur predominantly in conversation, where they account for nearly 45% of all lexical verbs, compared to only 11% in academic prose. This distribution reflects the interactive and real-time nature of spoken discourse and the greater lexical diversity of written registers.

Beyond these high-frequency items, **common lexical verbs** (occurring at least 50 times per million words) also show systematic variation across semantic domains. Activity verbs constitute the largest proportion (approximately 49%), followed by mental verbs (19%) and communication verbs (13%), highlighting the prominence of activity verbs in English. The distribution varies across registers. In conversation, activity, mental and communicative verbs are particularly frequent, aligning with its function of reporting actions and events, expressing thoughts and feelings, and presenting speech events. Conversely, academic prose makes extensive use of activity, causative and occurrence verbs, indicating its focus on reporting relationships among entities. News reportage frequently uses activity and existence verbs, consistent with its emphasis on event reporting and state description (Biber et al., 2021, pp. 363–371).

In addition to semantic variation, common lexical verbs (occurring at least 300 times per million words in at least one register) exhibit differences in valency patterns, including transitive, intransitive, and copular constructions. These patterns shape the range and type of elements that co-occur with verbs. Overall, 47% of common verbs alternate between transitive and intransitive patterns, while 36% occur exclusively with transitive patterns. Most common verbs take multiple valency patterns, and even those with similar patterning may differ markedly in their actual usage (Biber et al., 2021, pp. 380-384). For example, the verb *show* displays considerable variation in its valency patterns across registers and is more register-dependent than most verbs examined in the study. In conversation and fiction, it commonly occurs in ditransitive patterns with an indirect object (e.g., *I'll show you*), reflecting interpersonal interaction. In contrast, news and academic prose show a higher proportion of patterns involving complement clauses (e.g., *they still show boardroom salaries growing faster*) (pp. 389-391). This variation suggests that verb-based co-occurrence patterns are shaped not only by lexical choice but also by underlying grammatical structures, highlighting the relationship between valency and co-occurrence.

2.3.2.2 Empirical findings of studies using MD analysis

At the most comprehensive end of the continuum, MD analysis (Biber, 1988) provides a framework for identifying patterns of co-occurring linguistic features across texts. In this approach, the normalized frequencies of a large set of features are subjected to factor analysis to “empirically analyse the ways in which linguistic features co-occur in texts and the ways in which registers vary with respect to those co-occurrence patterns”. The resulting patterns are interpreted as *dimensions* of variation, each characterised by “its own set of co-occurring features that reflect different functional underpinnings” (Biber, 2019, pp. 49-50).

The original MD analysis: Biber's model

Biber's (1988) original MD analysis examined a broad range of spoken and written English registers. The written data were drawn primarily from the Lancaster-Oslo/Bergen (LOB) Corpus and covered 15 written registers, including "press reportage, editorials, press reviews, religion, popular lore, skills and hobbies, biographies and essays, official documents, learned writings, fiction (including general, mystery, adventure, science, and romance), and humor". The spoken data were from the London-Lund Corpus and comprised six registers: "private conversations, public conversations (including interviews and panel discussions), telephone conversations, radio broadcasts, spontaneous speeches, and prepared speeches" (p. 66). On the basis of 67 linguistic features, the analysis identified seven dimensions of variation.

Biber (2014) later provides a concise synthesis of the original MD model, summarising the major linguistic features and functional interpretations associated with each dimension. Table 5 draws directly on this summary and presents the major linguistic features co-occurring on Dimensions 1-5 (Biber, 2014, p. 11). Together, these dimensions demonstrate that English registers vary along continuous functional parameters, including "interactiveness, involvement and personal stance, production circumstances, informational density, informational elaboration, narrative purposes, situated reference, persuasiveness or argumentation, and impersonal presentation of information" (Biber, 2014, p. 15).

Table 5

Summary of the major linguistic features co-occurring on Dimensions 1–5 from the 1988 MD analysis of register variation

Dimension 1	Positive features
Involved vs. Informational Production	mental (private) verbs, 'that' complementizer deletion, contractions, present tense verbs, 'wh'-questions, 1st and 2nd person pronouns, pronoun 'it', indefinite pronouns, 'do' as pro-verb, demonstrative pronouns, emphatics, hedges, amplifiers, discourse particles, causative subordination, sentence relatives, WH-clauses
	Negative features
	nouns, long words, prepositions, type/token ratio, attributive adjectives
Dimension 2	Positive features
Narrative vs. Non-narrative Discourse	past tense verbs, 3rd person pronouns, perfect aspect verbs, communication verbs
	Negative features
	present tense verbs, attributive adjectives
Dimension 3	Positive features
Situation-dependent vs. Elaborated Reference	time adverbials, place adverbials, other adverbs
	Negative features
	WH-relative clauses (+subject gaps, object gaps), phrasal coordination, nominalizations
Dimension 4	Positive features
Overt Expression of Argumentation	prediction modals, necessity modals, possibility modals, suasive verbs, conditional subordination, split auxiliaries
Dimension 5	Positive features
Abstract/Impersonal Style	conjuncts, agentless passives, 'by'-passives, past participial adverbial clauses, past participial postnominal clauses, other adverbial subordinators

Among these, Dimension 1 provides the most comprehensive account of register variation (Biber, 2014). Features that load positively on this dimension characterise discourse with “interactional, affective, involved purposes, associated with strict real-time production and comprehension constraints”, whereas negative loadings characterise discourse with “highly informational purposes, which is carefully crafted and highly edited” (Biber, 1988, p. 115). Dimension 2 differentiates discourse with “primary narrative purposes” from discourse with “non-narrative purposes, such as expository or descriptive” (Biber, 1988, p. 115). Dimension 3 captures the contrast between “explicit, endophoric, text-dependent reference” and “situation-dependent reference”, which relies on the shared context of addresser and addressee (Biber, 1988, p. 110), with written texts typically favouring the former.

Dimension 4 is associated with “the speaker’s expression of own point of view or with argumentative styles intended to persuade the addressee” (Biber, 1988, p. 115), which is particularly salient in persuasive discourse domains like editorials and parliamentary debates (Huang & Ren, 2020; Kruger et al., 2019). This reflects the persuasive nature of parliamentary debates, where speakers intend to persuade listeners of a particular viewpoint through interactions characterized by confrontational, competitive, cooperative, and performative elements (Ilie, 2015). Dimension 5 differentiates between highly abstract and technical informational discourse and non-abstract, colloquial discourse (Biber, 1988).

For the four registers examined in the present study—conversation, parliamentary debates, news reports, and academic prose—these dimensions can be used to position the registers relative to one another. Table 6 below presents the rankings of these registers based on their mean dimension scores, situating them along the five parameters of variation adapted from Biber (1988, pp. 128-152).

Table 6
Rankings of Registers in Each Dimension

Score	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Positive	telephone conversations	spontaneous speeches	press reviews	editorials	academic prose
	face-to-face conversations	prepared speeches	academic prose	telephone conversations	press reviews
	spontaneous speeches	press reportage	editorials	prepared speeches	press reportage
	prepared speeches		spontaneous speeches prepared speeches	spontaneous speeches	editorials
Negative	editorials	face-to-face conversations	press reportage	face-to-face conversations	prepared speeches
	press reviews	editorials	face-to-face conversations	academic prose	spontaneous speeches
	academic prose	press reviews	telephone conversations	press reportage	face-to-face conversations
	press reportage	telephone conversations academic prose		press reviews	telephone conversations

Subsequent MD studies

Subsequent studies have applied the MD approach to diverse discourse domains, including conversation (e.g., Biber, 2004), parliamentary debates (e.g., Kruger et al., 2019), academic prose (e.g., Jin, 2021) and editorials (e.g., Huang & Ren, 2020). Across these dimensions, Dimensions 1 and 2 have been found to emerge repeatedly across MD studies, corresponding respectively to “the opposition between clausal/oral discourse vs. phrasal/literate discourse” and “the opposition between narrative vs. non-narrative discourse” (Biber, 2014, p. 7).

At the same time, other dimensions are more register-specific. For instance, Biber (2004) identified a dimension of *Stance vs. Context-Focused Discourse* in conversation text types, while Huang and Ren (2020) proposed *Expression of Institutional Stance* for editorials in China Daily and the New York Times. In specialized academic discourse, Jin (2021) developed a set of dimensions tailored to research article discussion sections in the engineering discipline. These include *Commitment to scientific propositions*, *Expressions of attitude*, and *Framing scientific claims*, which refine broader categories of stance and argumentation, as well as the dimension *Information production*. Additionally, *the absence of supporting literature or negation of scientific propositions* has no direct equivalent in Biber’s model. These findings indicate that, while some dimensions reflect general parameters of register variation, others capture functional distinctions specific to particular discourse domains. Table 7 summarises these specialized dimensions and the linguistic features associated with them.

Table 7

Specialized dimensions in selected MD studies

Dimension of Stance vs. context-focused discourse in Biber (2004)	Features with positive loadings: that-deletions, mental verbs, factual/mental verb + that-clause, likelihood/mental verb + that-clause, likelihood adverbs, adverbial clauses, general hedges, factual adverbs Features with negative loadings: nouns, WH-questions
Dimension of Expression of Institutional Stance in Huang and Ren (2020)	Features with positive loadings: article + adj. + noun, other adverbial subordinators, adj. + noun + verb, conjuncts, emphatics, split auxiliary Features with negative loadings: other nouns, word length, 3rd person pronouns, article + noun + and
Dimension of The absence of supporting literature or negation of scientific propositions in Jin (2021)	Features with positive loadings: negation, existential there, to-infinitives Features with negative loadings: prepositional phrases

Taken together, the conceptualisation of register, its analytical framework, and empirical findings from corpus-based studies of register variation demonstrate that variation in language use is systematic, functional, and closely shaped by situational and communicative factors. Both feature-based analyses and multidimensional approaches further reveal that linguistic features tend to co-occur in patterned ways that reflect communicative purposes and discourse functions across registers, providing significant insights into the distribution and functional interpretation of individual features as well as broader dimensions of variation.

From the perspective of valency theory and corpus pattern analysis, however, verbs occupy a central role in structuring lexico-grammatical patterns, as they determine the selection and organisation of their co-occurring elements. While studies of register variation have documented

systematic differences in verb-related features—such as frequency, semantic domains, and valency patterns (Biber et al., 2021)—these findings are typically based on analyses of individual features or abstract dimensions. As a result, they provide only a partial account of how verb-collocate relationships are realised across registers. In addition, the relationship between such patterns and the communicative purposes and discourse functions associated with particular registers has not been systematically explored. Extending this perspective, the following section shifts the focus to phraseological patterning across registers, where recurrent strings and strongly associated word combinations offer a more integrated perspective on the distributional and functional variation of verb collocations.

2.4 Phraseological patterning across registers: recurrence and co-occurrence

Drawing on the preceding discussions of phraseology and register variation, this section examines two types of phraseological patterning across registers. It first reviews research on recurrence, with a focus on lexical bundles (particularly verb phrase-based bundles), and then turns to co-occurrence, focusing on relational collocations involving verbs.

2.4.1 Recurrence: lexical bundles

Research on recurrence in register studies has primarily centred on lexical bundles, as developed in the work of Biber and his colleagues. These bundles are characterised by high frequency, semantic transparency, lack of perceptual salience, and structural incompleteness, distinguishing them from more fixed or idiomatic phraseological units (Biber, 2009).

Within this framework, lexical bundles are not treated merely as frequent sequences, but as “lexical building blocks” widely used across speakers and situations (Biber et al., 2021, p. 983). Structurally, lexical bundles can be grouped into three broad categories: bundles incorporating verb phrase fragments (e.g., *you don’t have to*), bundles incorporating dependent clause fragments (e.g., *I want you to*), and bundles incorporating noun phrase or prepositional phrase fragments (e.g., *one of the things*). Functionally, they are broadly grouped into three main functional categories: stance expressions (e.g., *I don’t know if*), discourse organizers (e.g., *what do you think*), and referential expressions (e.g., *that’s one of the*) (Biber et al., 2004). Crucially, these bundle types vary systematically across registers, thereby contributing to the differentiation of discourse types and linking phraseological patterning to register variation (Conrad & Biber, 2005).

The primary objective of this line of research is to “identify the most common recurrent sequences of words and to determine the extent to which these sequences can be interpreted as building blocks of discourse” (Conrad & Biber, 2005, p. 58). Accordingly, studies typically investigate whether particular sequences occur with high frequency, whether registers are characterized by distinct sets of such sequences, and how these bundles perform discourse functions and contribute to the communicative repertoires of speakers and writers (Conrad & Biber, 2005). Empirical findings consistently show that analyses of recurrent word sequences constitute a useful tool in register studies and corpus-linguistic analyses (Crossley & Louwerse, 2007), as evidenced in research on conversation, parliamentary debates, newspaper discourse, and academic prose (Conrad & Biber, 2005; Granger, 2014; Gaspari, 2013; Jablonkai, 2009; Wu, 2021). Table 8 presents a comparative overview of these studies, while the discussion below focuses on verb phrase-based bundles and the distribution of discourse functions across bundle types.

Table 8

Summary of Major Findings on Lexical Bundles across Registers

Study	Register(s)	Bundle scope	Analytical focus	Frequency and structural characteristics	Functional patterns
Conrad & Biber (2005)	Conversation vs. academic prose	3- and 4-word bundles	Frequency, structure, function	Conversation: ~90% contain VP elements; academic prose: ~60% are NP/PP-based.	Conversation: stance bundles dominant; academic prose: referential bundles dominant; discourse organisers relatively infrequent in both.
Granger (2014)	Parliamentary debates vs. newspaper editorials	Lexical bundles, specifically stems (subject + verb)	Types and tokens; keyness	Parliamentary debates are more formulaic: almost three times as many stem types and tokens as editorials, with a lower type/token ratio indicating greater repetition. A substantial shared repertoire of stem tokens (29%) exists, yet 60% of stems occur exclusively in parliamentary debates.	Parliamentary debates: first-person stems; directive stance. Editorials: impersonal <i>it/there is (are)</i> structures; a more reflective stance.
Wu (2021)	EU parliamentary debates (European Parliament plenary sessions)	4-word bundles	Structure and function	NP/PP-based: 64%; VP-based: 20%; dependent clause bundles: 15%.	Subject-specific bundles: 50%; referential: 24%; stance: 20%; discourse organisers: 6%; oral features remain visible in desire bundles, discourse markers, and metadiscourse bundles.
Jablonkai (2009)	Online EU news discourse	4-word bundles	Frequency, structure, function	Top 40 bundles show slight differences from the general BNC but considerable differences from the academic BNC. NP/PP-based: ~80%; dependent clause bundles: ~10%; VP-based bundles: <10%.	Subject-specific bundles most frequent; referential bundles also common; stance bundles and discourse organisers comparatively infrequent; time-referential bundles are the most common referential subtype.
Gaspari (2013)	Online news reports (Reuters vs. UPI)	4-word bundles	Quantitative and qualitative comparison	UPI contains almost twice as many lexical bundles as Reuters; after excluding proper-name bundles, Reuters uses bundles about twice as frequently as UPI per million words. Frequent bundles include formulaic reporting expressions.	Study focuses primarily on quantitative and structural comparison between wire services; functional patterns not analysed separately.

Note. BNC = British National Corpus; NP = noun phrase; PP = prepositional phrase; VP = verb phrase; EU = European Union. All frequency figures are approximate values reported in the original studies.

Verb phrase–based bundles across registers

Building on this overview, particular attention has been given to verb phrase-based (VP-based) bundles, which provide a more direct perspective on the use of verbs within lexico-grammatical patterns across registers.

In their comparison of **conversation** and **academic prose**, Conrad and Biber (2005) found that conversation exhibited a higher overall frequency of three- and four-word bundles, as well as greater textual coverage than academic prose. VP-based bundles, together with dependent clause fragments, were more common in spoken discourse. Specifically, they occurred as parts of declarative clauses or questions, with approximately 90% containing elements of a verb phrase. Functionally, stance bundles were dominant in conversation, whereas referential bundles were more frequent in academic prose, while discourse organizers remained relatively infrequent in both registers. These patterns reflect the contrast between the interpersonal orientation of conversation and the informational focus of academic discourse.

Extending this comparison to other registers, Granger (2014) examined stems (defined as sequences that contain a subject and a verb such as *it is important that*) in **parliamentary debates** and **newspaper editorials**. The results showed that parliamentary debates were more formulaic, displaying substantially higher numbers of stem types and tokens and a lower type/token ratio than editorials, indicating greater repetition. Moreover, register-specific patterns emerged: parliamentary debates favoured stems containing first-person pronouns (mainly singular forms), reflecting the ritualized nature and special illocutionary force (e.g., *Mr President I welcome*). Stance-expressing stems were also common, especially those conveying directive stance. Conversely, editorials relied more on impersonal structures with the pronoun *it* and existential structures with *there is (are)*, signalling a more reflective stance.

In a similar vein, Wu (2021) analysed four-word bundles in European Parliament (EP) plenary **parliamentary discourse** and found that VP-based bundles accounted for approximately 20% of all bundles. Despite this, the overall structural distribution resembled that of academic discourse, pointing to the informational orientation and relatively literate nature of parliamentary speech. As Wu (2021) noted, “70% of the speeches are either read-out or a mix between read-out and impromptu delivery” (p. 61). Functionally, EP parliamentary debates were largely information-oriented, reflected in the dominance of subject-specific bundles (50%). These included references to EU organizations/institutions (e.g., *Committee of the Regions*), EU documents (e.g., *the Treaty on European*), Codes (e.g., *# of the EC*) and Others (e.g., *the principle of subsidiarity*). At the same time, parliamentary debates preserved salient oral features. This was indicated by the presence of desire bundles (e.g., *if you want to*), discourse markers, and meta-discourse bundles (e.g., *go back to the*), alongside the absence of text-deixis bundles (e.g., *as shown in figure*). Taken as a whole, this distribution highlights the hybrid spoken-written nature of parliamentary discourse.

In **news discourse**, Jablonkai (2009) reported that VP-based bundles unexpectedly accounted for less than 10% of bundles in online EU news texts, a relatively low proportion given that news texts typically report on events and actions and are written for the general public. One possible explanation is that diverse verbs are used to report events, reducing the recurrence of specific verb phrases. In terms of overall bundle distribution, online EU news discourse showed only slight differences from the general BNC corpus, but considerable differences from its academic component.

Functionally, subject-specific bundles dominated (60%), encompassing a range of subtypes, particularly organization-related bundles, alongside moderate use of person-, event-, and other-related bundles. Referential bundles were also frequent, with time-referential expressions being the most common subtype, reflecting the need for temporal and spatial specification in event

reporting. Stance bundles and discourse organizers, by contrast, were comparatively rare. When present, **stance bundles** typically expressed desire, obligation, and intention. The relatively diverse use of such attitude bundles may reflect the tendency of news reports to incorporate perspectives of multiple participants involved in a single event. The limited use of **discourse organizers** may indicate a preference for alternative linguistic means or lexical items to signal structure and organize discourse.

Taken together, the studies reviewed above show that as recurrent “building blocks of typical discourse within the register” (Conrad & Biber, 2005, p. 69), lexical bundles display systematic differences in frequency, structural characteristics, and discourse functions, reflecting variation in communicative purposes and discourse organisation across registers.

Methodological limitations of recurrence analysis

Despite their usefulness for identifying recurrent phraseological patterns, the extraction of lexical bundles is sensitive to methodological decisions concerning bundle length and frequency thresholds (Gaspari, 2013). Overlapping sequences such as *the end of the* and *by the end of*, together with shorter units such as *the end of*, illustrate the difficulty of delimiting bundles and establishing principled extraction criteria. More broadly, as lexical bundles are identified as recurrent contiguous sequences, such procedures may privilege particular surface strings rather than capture broader phraseological patterns that are realised in structurally more variable forms.

A related methodological strand has examined the extent to which n-grams can discriminate among registers. Gries et al. (2011) show that n-gram distributions can recover meaningful register structure, including a robust spoken-written distinction, with 3-grams displaying the strongest discriminatory power, while also revealing interpretable relationships among finer-grained sub-registers. Crossley and Louwerse (2007) demonstrate that bigrams are useful for register classification and for capturing syntactic and discourse features. However, they also acknowledge that “a bigram approach to register classification has limitations” (p. 475): although it works well in distinguishing disparate registers, it is less effective in discriminating among closely related registers. This limitation was evident in the London-Lund Corpus, where several spoken registers—including radio and television broadcasts, casual face-to-face and telephone conversations, radio and telephone interviews, and spontaneous and prepared speeches—were positively associated with only one register dimension. Goulart et al. (2020) similarly note that n-gram methods do not consistently “predict specific registers within those modes that differ in terms of communicative purpose (i.e., specific sub-registers within modes)” (p. 4).

Taken together, these limitations indicate that recurrence, as captured through n-gram analysis, provides only a partial account of phraseological patterning across registers. Lexical bundle analysis is constrained to recurrent contiguous strings and is therefore less suited to identifying lexical relationships that allow structural variation or non-adjacent co-occurrence. In addition, although n-gram distributions can distinguish broad register categories, their discriminatory power may be more limited for closely related registers. These limitations motivate a complementary focus on co-occurrence.

2.4.2 Co-occurrence: statistical collocations

In comparison with the extensive body of research on lexical bundles, co-occurrence—particularly statistical collocations across registers—has received relatively limited and less systematic attention. Co-occurrence analysis captures a different aspect of phraseological patterning: the statistical association between lexical items within a defined span or syntactic structure. This perspective is especially relevant for examining verb collocations, which do not necessarily form fixed sequences. By identifying such associations through statistical measures, co-occurrence-

based approaches complement recurrence-based approaches and offer a more flexible account of how phraseological patterns vary across registers.

Collocational variation across registers: frequency and association strength

Spina and Tanganelli (2012) investigated collocations across six Italian registers, including three written registers (academic texts, literary prose, school essays) and three spoken registers (dialogic speech, monologic speech, film dialogues), focusing on four productive POS sequences in Italian (i.e., verb-noun, noun-noun, noun-preposition-noun and noun-adjective). Based on the frequency distribution, their findings showed that collocational frequency differed across registers and could contribute to their differentiation, although frequency alone was not sufficient to explain the observed differences. To provide a more fine-grained account, they employed lexical gravity as the association measure and subjected these values to cluster analysis, which grouped registers according to similarities in the type-frequency patterns of their collocational constituents rather than raw frequency alone. This analysis grouped literary texts and film dialogues together on the basis of their shared narrative function, while academic prose formed a distinct group from the other registers, which was characterised by a higher proportion of noun collocations and fewer verb collocations, a pattern consistent with its informational orientation.

Adopting a broader perspective on formulaic language (which was used in that study as a cover term for the phenomenon of formulacity), Vilkaitė (2016) examined the overall **frequency** distribution of collocations as one category, alongside lexical bundles, phrasal verbs and idiomatic phrases, in the BNC Baby corpus, which includes academic prose, fiction, newspaper language, and spoken conversation. Collocations were defined as two- or three-word sequences occurring either adjacently (e.g., *hard work*) or within a limited span allowing one intervening word (e.g., *take a break*), and were identified using combined frequency (5) and MI thresholds. Methodologically, all bigrams and trigrams were first extracted and then filtered based on MI scores to identify candidates exhibiting mutual expectancy (MI 3), with sequences consisting solely of function words or containing proper names excluded. The resulting dataset comprised 9,913 two-word and 3,262 three-word collocations, including nonadjacent combinations (e.g., *taken into account*), binomials (e.g., *black and white*), and extended collocations (e.g., *indexed sequential file*) (p. 36).

The results showed that collocations accounted for 7.52% of the corpus overall, with clear variation across registers: they were most frequent in academic prose (97,406; 9.72%), followed by newspaper language (72,261; 7.52%), and less common in spoken conversation (69,102; 6.82%) and fiction (61,375; 6.04%). This distribution suggests a gradient across registers, with more informational registers favouring higher collocational density than more interactive or narrative ones. Further analysis indicated that these differences in frequency were statistically significant across both registers and language modes, although variation was more strongly associated with the spoken-written distinction than with finer register categories (p. 40).

Beyond frequency, the strength of association between lexical items represents a further dimension of variation. Gablasova et al. (2017) investigated three types of collocations—verb + complementation (*make + sure/decision/point*), adjective + noun (*human + beings/right/nature*), and adverb + adjective (*vitally/very/really + important*)—across the BNC as a whole and its constituent sub-corpora (academic writing, newspapers, fiction, formal and informal speech), operationalising collocational strength through MI score, t-score, and log Dice.

The results showed that t-score values varied substantially across corpora because of its sensitivity to corpus size, whereas MI scores and Log Dice were more stable and were therefore more suitable for cross-corpus comparison. The study also demonstrated systematic variation in collocational strength across registers and across the written/spoken divide, suggesting that

analyses based on large, heterogeneous corpora may hide important register- and mode-related differences.

These findings also point to the limitations of AM-based analysis. The authors therefore argue for combining frequency-based and phraseological approaches in future research in order to capture a broader range of collocational properties. AMs alone provide only a partial account of collocational knowledge, as frequency-based accounts are largely blind to linguistic properties such as “semantic unity and fixedness of form” (p. 173). Collocations identified by high or low association scores should thus be further examined through structural and semantic analysis to obtain a clearer picture of the collocational patterns associated with particular groups of speakers or types of word co-selection.

Verb-based case studies

Such an integrative perspective motivates closer examination of verb-centred collocational patterns across registers. In this regard, case studies of common or high-frequency verbs constitute an important strand of collocation research, as they aim to reveal the lexico-grammatical patterns of these verbs and their corresponding functions. However, relatively few corpus-based studies have systematically examined how the use of high-frequency verbs varies across registers, particularly in comparisons that span both spoken and written modes. Notable exceptions include Biber’s aforementioned work on common lexical verbs. Among high-frequency verbs, *make* and *take* have received especially extensive attention.

A representative example is the study by Conrad and Biber (2009), which focused specifically on the verbs *make* and *take*. By categorizing and contrasting their noun collocates across conversation and academic prose, Conrad and Biber (2009) demonstrated that both verbs were extremely common in both registers and were predominantly used in a range of idiomatic meanings rather than in their literal senses. The collocates of both verbs were grouped into semantic categories and then compared between the two registers. For *make*, ten semantic groups were identified in conversation, compared to five in academic prose, where collocates tended to be more abstract (e.g., *make reference to*). Similarly, *take* occurred with eleven semantic groups in conversation, and seven in academic prose. In written registers, typical collocates such as *action*, *position*, and *advantage* reflected more specialized meanings (e.g., initiating actions or adopting positions), often followed by a certain preposition such as *take advantage of*. Overall, the findings highlight that noun collocates of high-frequency verbs are important for examining register variation, particularly in relation to differences in semantic range and lexico-grammatical patterns.

Extending this line of inquiry, Han and Shin (2015) investigated the same two verbs across two academic sub-registers—classroom teaching (lectures) and textbooks—offering a comprehensive analysis of their lexico-grammatical and functional uses. They identified nine functional categories for lexico-grammatical patterns (verb + complement) of each verb. For *make*, these included Produce (e.g., *make salad*), Delexical (e.g., *make a decision*), Causative (e.g., *make sure that*), Money (e.g., *make money*), Link Verb (e.g., *make a teacher*), Make it (idiomatic), Phrasal, Idiomatic (e.g., *make one’s way*), and Others. For *take*, the categories included Move (e.g., *take people somewhere*), Delexical (e.g., *take an action*), Consider (e.g., *take account of*), Transport (e.g., *take a bus*), Time (e.g., *take a minute*), Get/Use (e.g., *take a number*), Phrasal, Idiomatic (e.g., *take an example*), and Others.

Their findings revealed systematic variation across the two sub-registers. *Take* was more frequent in less formal, spoken-like contexts, whereas *make* was more characteristic of formal, written contexts. The Delexical, Causative, and Produce categories of *make*, as well as the Delexical, Get/Use, and Move categories of *take*, occurred significantly more often in lectures

than in textbooks, while phrasal and idiomatic uses showed no significant variation. These results confirm earlier findings on spoken-written differences, demonstrating that register variation operates not only at the level of frequency but also in the distribution of functional and lexico-grammatical patterns.

In addition to frequency analysis, the study examined the 20 most frequent collocates of each verb and classified them according to Biber's (2006) semantic categories. The results showed clear register differences in collocational patterns: verbs and nouns were more prominent as collocates in lectures, whereas adjectives were more prominent in textbooks. Although the specific collocates of individual verbs varied across the two registers, stance-related words occurred commonly as collocates in both registers. In the written register, the verb *make* with adjective collocates (e.g., *make a difficult decision*) tended to express evaluative meanings. The authors attributed this pattern to the composition of their textbook corpus, which included Business and Social Science textbooks in which personal judgements and arguments were relatively prominent, despite textbook language being generally characterised as relatively impersonal.

Adopting a different analytical framework, Alqarni (2019) applied the Pattern Dictionary of English Verbs (PDEV) approach to the verb *suggest*. Based on data from the BNC, the study first constructed a PDEV-style entry for *suggest* and subsequently examined its distribution in the Louvain Corpus of Native English Essays (LOCNESS). All 51 instances in the native-speaker corpus could be accounted for by the established patterns, and no atypical or new prototypical pattern was identified, despite the two corpora representing different registers. What differed, however, was the relative frequency of the patterns, as illustrated in Table 9 (adapted from Alqarni, 2019, pp. 172–175).

From the perspective of the present study, this finding is noteworthy, as it suggests that register variation in verb use is more likely to be reflected in differences in the distribution of existing patterns than in the emergence of new ones. Although this analysis is limited to a single verb and does not further explore variation in semantic types, contextual roles or lexical sets within individual patterns, it nonetheless underscores the value of extending verb-based analyses to a wider range of verbs and datasets.

Table 9

PDEV Entry for Suggest with Frequencies of Patterns in the Two Corpora

No.	Pattern	% frequency in LOCNESS	% frequency in BNC sample
1	Human suggests Action or Plan or Proposition or <i>that</i> -clause or <i>ing</i> -clause Quote	19.60	23
2	Human suggests <i>that</i> -clause	33.33	25
3	Eventuality 1 suggests Eventuality 2 or <i>that</i> -clause	39.21	52
4	Human suggests Name	0	0

On the whole, the literature reviewed in this section demonstrates that phraseological patterning varies systematically across registers. However, several limitations remain. First, research on recurrence has largely relied on n-gram approaches, which prioritise contiguous sequences and are less effective in capturing non-adjacent word combinations. Second, studies of co-occurrence are comparatively underrepresented in the literature, with verb-based analyses constituting a major line of research within this area. While these provide detailed insights, they are typically limited to a small set of high-frequency verbs and a narrow range of registers, and do not fully incorporate structural, semantic, and functional perspectives, thereby making it difficult

to generalise the findings across broader contexts of language use. Collectively, these gaps point to the need for a broader and more systematic investigation of collocations across registers. The next section situates the present study by outlining the research gap, highlighting its significance, and presenting research questions.

2.5 Positioning of the present study

The preceding sections reveal several unresolved issues in the existing literature. Although collocation research has made important methodological advances, it has yet to provide a comprehensive account of how verb-based co-occurrence patterns vary across communicative contexts in terms of frequency, distribution, semantic profile, and discourse function. Meanwhile, although register studies have offered systematic descriptions of linguistic variation, they have tended to treat verbs as individual linguistic features rather than as elements of broader phraseological patterning. As a result, the role of verb-centred collocational patterns remains underexplored.

This gap is particularly important for two reasons. First, the specificities of different registers and their influence on the nature of observed linguistic phenomena have not been fully examined (Dupont & Zufferey, 2017; Gablasova et al., 2017). Second, understanding how different linguistic contexts shape word co-occurrence is central to explaining how language becomes formulaic (Ellis et al., 2015). However, existing research on phraseological variation across registers has focused more extensively on recurrent contiguous sequences than on co-occurrence patterns. Studies that do examine verb collocations are often confined to small-scale analyses of selected high-frequency verbs, and therefore provide only a partial account of how verb collocations vary across structural, semantic, and functional dimensions.

To address these limitations, the present study investigates verb collocations across four registers: informal conversation, parliamentary debates, news articles, and academic journal articles. It operationalises word combinations as statistical collocations and, more precisely, as relational collocations identified within grammatical relations. The analysis focuses on high-frequency lexical verbs shared across the four registers and examines how their collocational patterns vary in terms of frequency, distribution, register sharing, semantic profile, and discourse function. Accordingly, the study is guided by the following three research questions:

1. To what extent does the use of verb collocations vary across registers?
2. How does register influence the frequency and distribution of verb collocations across registers?
3. In what ways do the collocational patterns of these verbs reflect the communicative purposes and discourse functions specific to each register?

3 Methodology

3.1 Introduction

This chapter outlines the methodological procedures used to investigate verb collocations across registers. To address the research questions, the study adopted a corpus-based approach focusing on 25 high-frequency lexical verbs shared across the four registers. The study first examined the grammatical relations in which these verbs occurred, and then analysed their relational collocations across grammatical relations and registers in terms of frequency, distribution, register sharing, and semantic profile. By combining quantitative distributional analysis with qualitative and semantic interpretation, the study explored how register influenced collocational patterns of these lexical verbs.

The chapter starts by introducing the corpora used in the study and explaining the rationale for their selection in **Section 3.2**. **Section 3.3** describes how the top 25 shared lexical verbs were identified for further analysis. **Section 3.4** details the extraction of candidate relational collocations from the Word Sketch output and the validation procedures used to prepare the dataset for analysis. Building on these procedures, **Section 3.5** presents the analytical procedures used to examine the collocational patterns of these lexical verbs across registers. Finally, a conclusion is provided in **Section 3.6**.

3.2 Corpora

The present study draws on four corpora representing distinct registers: the spoken component of the British National Corpus 2014 for conversation, the ParlaMint 2.1 corpus of English parliamentary debates for parliamentary debates, the UK component of the SiBol corpus of English Broadsheet Newspapers 1993–2021 for news reporting, and the academic component of the British National Corpus XML Edition for academic writing. The corresponding corpus resources are cited as BNC2014, spoken part (Love et al., 2017), ParlaMint 2.1 (Erjavec et al., 2021), SiBol, compiled by the SiBol research group and accessed through Sketch Engine (a web-based corpus analysis platform; Kilgariff et al., 2014), and the BNC XML Edition (The BNC Consortium, 2007).

The selection of these registers is theoretically motivated by their contrasting situational characteristics, including participants, relations among participants, channel, processing circumstances, setting, communicative purposes, and topics (see Table 3). On a practical level, the corpora were chosen for their availability in Sketch Engine and for their methodological comparability, as all four corpora are annotated using the same tagset and Word Sketch Grammar. Table 10 provides an overview of these corpora, including their size, tagset, Word Sketch Grammar, representativeness, and corpus description.

Table 10*Overview of (sub)corpora used*

	British National Corpus, academic part	English Broadsheet Newspapers 1993–2021 (SiBol), UK component	English parliamentary debates (ParlaMint 2.1)	British National Corpus 2014 (BNC2014), spoken part
Tokens (running words)	17,625,626	783,648,132	111,980,128	11,832,933
Word Types	15,083,037	666,659,805	100,616,051	10,495,185
Tagset	English TreeTagger PoS tagset with Sketch Engine modifications	English TreeTagger PoS tagset with Sketch Engine modifications	English TreeTagger PoS tagset with Sketch Engine modifications	English TreeTagger PoS tagset with Sketch Engine modifications
Word Sketch Grammar	English Penn Treebank tagset, version 3.3	English Penn Treebank tagset, version 3.3	English Penn Treebank tagset, version 3.3	English Penn Treebank tagset, version 3.3
Mode and Register	Written, academic writing	Written, news	Spoken, parliamentary debates	Spoken, conversation
Corpus Description	Academic proeses from disciplines of humanities, medicine, natural sciences, politics law education, social and behavioural sciences, and technology computing engineering	Texts come from newspapers of the years 1993–2021: (1) Broadsheets: The Times, The Guardian, The Daily Telegraph, Sunday Times, Sunday Telegraph; (2) Tabloids: The Sun, Daily Mirror, Daily Mail, Daily Express; (3) Free papers: Metro, Evening Standard	British parliamentary debates from 2015 to 2021	Informal conversations recorded among friends and family members across the UK (2012–2016)

3.3 Selection of target lexical verbs

To establish a core set of cross-register lexical verbs, a comparative frequency analysis was conducted across the four corpora. In each corpus, the 1,000 most frequent verbs were extracted using the Wordlist function in Sketch Engine. The part-of-speech setting was restricted to *verb*. This procedure retrieved lexical verbs as well as primary auxiliary verbs *be*, *have*, and *do*. Modal verbs, such as *might*, were not included because they are tagged separately as MD in the Word Sketch Grammar. Since the four corpora differ considerably in size, the downloaded Wordlist outputs were ranked by Frequency per million tokens, a relative frequency value ($= \text{raw frequency} / \text{corpus size} \times 1,000,000$) provided by Sketch Engine, rather than by raw frequency.

The four Wordlist outputs were then imported into R for further processing. An intersection operation was performed to identify verbs that occurred in all four corpora. To rank these shared verbs, Average Rank was adopted as the primary sorting criterion. In each corpus, verbs were ranked according to their relative frequency, with rank 1 assigned to the most frequent verb. The four rank values obtained for each shared verb were then averaged to produce an overall rank score. A lower Average Rank therefore indicates that a verb is consistently frequent across the four registers. In this respect, Average Rank was considered more suitable than Mean Relative Frequency because the latter may be disproportionately affected by high-frequency outliers in a single register. Finally, 494 verbs were identified across all four corpora (see **Appendix A**). After the primary verbs *be*, *have* and *do* were manually excluded, the top 25 verbs ranked by Average Rank were selected as target lexical verbs. These verbs are listed in Table 11, together with their Average Rank and relative frequency in each register.

3.4 Extraction and validation of relational collocations

The Word Sketch function in Sketch Engine was used to extract candidate verb collocations. As illustrated in Figure 2 below, this tool provides a one-page summary of the typical grammatical and collocational behaviour of the node word. Collocates are grouped according to grammatical relations, within which the extracted word combinations are treated as relational collocations in the present study. For each collocate, the output also provides quantitative information, including the frequency of the node word, the frequency of each collocate within a given grammatical relation, and the association strength measured by logDice. In practical terms, logDice is useful for comparing collocations across corpora of different sizes, and it highlights strongly associated word combinations without giving excessive weight to rare items (Gablasova et al., 2017).

The grammatical relations displayed in the Word Sketch output are defined by the Sketch Grammar, a set of predefined lexico-grammatical rules used to detect collocates and classify their relationship to the node word (Kilgariff & Tugwell, 2002). In the English Word Sketch Grammar applied to the Sketch Engine corpora used in this study, subject collocates are typically noun-tagged words occurring in subject-like positions, such as before the verb, before a *have* + verb construction, or after *by* in passive constructions. Object collocates are noun-tagged words occurring in object-like positions, including after the verb, before a past-participle verb form, and before a *be* + past-participle construction. Modifier collocates are typically adverb-tagged words occurring before or after the verb, while complement collocates are adjective-tagged words occurring after the verb, either following a noun phrase or a personal pronoun.

Table 11
Top 25 Shared Lexical Verbs Ranked by Average Rank

No.	Verb	Average rank	CONV	PARL	NEWS	ACAD
1	say	7.00	4,714.47	2,704.14	4,552.80	780.23
2	make	7.25	1,542.31	3,156.26	1,921.26	1,735.43
3	take	9.75	1,470.22	2,000.13	1,586.89	1,175.33
4	see	10.00	2,450.03	1,196.52	1,135.00	1,283.30
5	go	10.50	6,961.17	1,621.24	1,738.68	509.94
6	know	11.00	7,793.17	1,279.24	929.21	619.78
7	give	13.00	822.70	1,585.53	977.91	1,253.46
8	come	13.75	2,119.25	1,199.98	1,183.82	528.89
9	think	16.25	5,977.55	1,137.62	901.99	414.40
10	use	17.00	1,300.52	704.42	724.61	1,407.67
11	get	17.75	8,591.78	1,314.10	1,673.84	290.94
12	need	17.75	1,264.27	1,629.18	677.57	471.08
13	work	21.75	904.26	1,189.22	669.88	437.43
14	find	22.25	671.09	534.03	816.11	864.65
15	look	24.50	1,680.99	962.71	787.90	296.78
16	leave	31.75	564.95	676.62	672.72	313.24
17	put	32.50	1,434.05	805.55	541.65	263.08
18	mean	33.00	2,191.26	458.97	371.78	412.98
19	want	34.25	2,050.29	1,242.28	830.64	201.69
20	pay	40.00	556.16	544.35	435.61	281.01
21	call	43.50	618.61	329.68	507.91	320.16
22	ask	44.00	440.80	750.08	454.99	245.38
23	show	44.00	223.28	335.89	515.09	864.71
24	try	47.00	826.42	442.78	496.43	222.57
25	bring	47.75	269.08	618.39	383.95	285.89

WORD SKETCH English parliamentary debates (ParlaMint 2.1) [Get m](#)

say as verb 302,810x ...

modifiers of "say"	objects of "say"	subjects of "say"	"say" and/or ...
1 rightly 1,880 10.1 ... rightly said	1 word 1,580 10.5 ... say a few words	1 Minister 13,314 11.0 ... Minister said	1 do 286 11.3 ... said and done
2 just 4,016 10.0 ... has just said	2 something 1,980 10.3 ... say something about	2 State 2,826 9.7 ... the Secretary of State said	2 write 26 9.0 ... written and said
3 before 1,182 9.8 ... I have said before	3 yesterday 852 9.9 ... said yesterday	3 gentleman 2,485 9.4 ... hon. Gentleman says	3 let 20 8.7 ... say , let us
4 simply 1,353 9.4 ... I simply say	4 today 1,560 9.9 ... said today	4 Friend 2,563 9.2 ... hon. Friend says	4 give 27 8.4 ... I have said , given
5 gently 766 9.3 ... I gently say to	5 thing 1,539 9.8 ... say two things	5 Lord 1,897 9.2 ... the noble Lord said	5 think 18 8.3 ... thought and said
6 also 3,274 9.0 ... also said that	6 anything 661 9.4 ... to say anything	6 Lords 1,827 9.1 ... noble Lords have said	6 intervene 13 8.1 ... to intervene and say
7 again 887 8.9 ... I say again that	7 week 683 9.3 ... said last week	7 friend 1,654 9.1 ... my noble friend said	7 imply 9 7.9 ... says , or implies
8 no 819 8.8 ... say no	8 nothing 621 9.1 ... say nothing	8 Secretary 1,653 9.1 ... Secretary said	8 come 21 7.9 ... come and say
9 already 1,719 8.8 ... I have already said	9 moment 378 8.8 ... I said a moment ago	9 baroness 1,273 8.8 ... the noble Baroness said	9 apologise 9 7.9 ... apologised and said that
10 clearly 777 8.7 ... said clearly	10 time 801 8.5 ... have said many times	10 other 1,296 8.8 ... others have said	10 stand 10 7.8 ... stand and say
11 repeatedly 496 8.6 ... have said repeatedly	11 bit 311 8.2 ... say a bit more about	11 Lady 1,337 8.8 ... hon. Lady says	11 get 9 7.3 ... say , get
12 previously 549 8.6 ... I have said previously	12 morning 205 8.1 ... said this morning that	12 member 1,494 8.7 ... Members have said	12 look 8 7.2 ... look and say

Figure 2: Word Sketch of *say*

Since the grammatical relations presented by Word Sketch are automatically generated, a small set of examples is provided in Table 12 to clarify how the major grammatical relations were interpreted in the present study. Most examples are drawn from the case-study verb *say* in **Section 4.3**; where a relation is less clearly illustrated by *say*, an example from *make* is provided. These examples show that the same node verb can occur in different grammatical environments, each of which provides a separate basis for identifying lexical collocates in specific syntactic relations. In this way, Word Sketch output provides an automatic, rule-based starting point for identifying relational collocations. However, because collocates are assigned to grammatical relations automatically, all candidate collocations required manual checking before being included in the final dataset.

Table 12

Illustrative Examples of Word Sketch Grammatical Relations

Grammatical relation	Example collocation	Concordance example
Subjects of X	<i>speaker + say</i>	The speaker says “the whole nation is quite intelligent anyway, as you can see”: the picture book is opened and the audience can see an ample, bare-breasted, black girl.
Objects of X	<i>say + word</i>	I should like to see them dare say a word against me.
Modifiers of X	<i>explicitly + say</i>	He explicitly says they do not constitute “knowledge” but what he calls “belief” or “opinion”.
X and/or . . .	<i>write and say</i>	Her first task will be to learn to write and say the numbers.
Infinitive objects of X	<i>say + represent</i>	A bundle of qualities is said to represent a “complete complex of compresence” if all its qualities are “compresent” and if nothing external to the bundle is compresent with all the qualities within it.
Complements of X	<i>make + clear</i>	The Directive does make it clear , however, that other types of contracts, such as those relating to employment, family law, company law and partnership, should be excluded from the Directive.

To improve the reliability of the dataset, two inclusion thresholds were applied before manual validation: a minimum raw frequency of 5 (following Vilkaitė, 2016) and a minimum logDice score of 5 (following Frankenberg-Garcia et al., 2019). Only collocates meeting both thresholds were retrieved for further inspection. Their concordance lines were then checked to determine whether each item functioned as an independent lexical collocate and whether it stood in the grammatical relation assigned by Word Sketch. Misclassified collocates were excluded or reassigned where appropriate. For multifunctional items, such as *well*, which may function as either a discourse marker or an adverb, classification was based on the item’s predominant function in context (see **Section 4.2.2** for examples). This validation process was documented in a separate spreadsheet recording retained relational collocations, excluded cases, and reasons for exclusion. The full validation record for the 25 lexical verbs is provided in **Appendix B**. The following five criteria guided the exclusion and reclassification of candidate collocates.

The first validation principle concerned the lexical status of the collocate. With reference to Biber et al. (2021, p. 64), only lexical words, which they define as nouns, verbs, adjectives, and adverbs, were retained as relational collocations, and items were included only when they functioned as independent lexical elements in a grammatical relation with the node verb. In accordance with this principle, function words were excluded, including determiners, pronouns, primary and modal auxiliaries, prepositions, adverbial particles (e.g., *around*, *home*, *past*, *through*), coordinators (e.g., *either* and *neither*), subordinators (e.g., *after* and *although/though*),

wh-words, existential *there*, the negator *not*, the infinitive marker *to*, and numerals (Biber et al., 2021, pp. 80, 97–98). Items functioning primarily as inserts were also removed, including interjections (e.g., *oh*), greetings and farewells (e.g., *alright*), discourse markers (e.g., *well*, *right*, *okay*), response elicitors (e.g., *look*), responses (e.g., *yeah*), hesitators (e.g., *er* and *erm*), thanks, the politeness marker *please*, apologies (e.g., *sorry*), and expletives (e.g., *Oh Jesus*) (Biber et al., 2021, pp. 97–98).

A second principle concerned modifier collocates. Although modifiers of verbs were expected to be adverbial, not all adverbs extracted by Word Sketch were treated as lexical modifiers. The validation followed five criteria:

- (1) General frequency adverbs and time adverbs were excluded, following Ackermann & Chen (2013, p. 239). Examples attested in the dataset include *daily*, *weekly*, *monthly*, *once*, *twice*, *always*, *usually*, *often*, *seldom*, *sometimes*, *hardly*, *rarely*, *never*, *now*, *then*, *today*, *yesterday*, *soon*, *later*, *already*, *yet*, *still*, *just*, *before*, *recently*, *since*, *ago*, *ever*, *again*, *far*, and *away*.
- (2) Place adverbs, such as *there*, *here*, *nowhere*, *somewhere*, *anywhere*, and *elsewhere*, were excluded.
- (3) Linking adverbials were excluded when their primary function was to connect clauses, sentences, or larger stretches of discourse rather than to modify the verb. These included additive, adversative, causal, and sequential adverbials such as *additionally*, *again*, *also*, *besides*, *further*, *furthermore*, *moreover*, *too*, *namely*, *likewise*, *however*, *though*, *yet*, *actually*, *conversely*, *instead*, *rather*, *anyhow*, *anyway*, *still*, *consequently*, *hence*, *naturally*, *so*, *therefore*, *thus*, *otherwise*, *afterwards*, *eventually*, *first(ly)*, *finally*, *last(ly)*, *next*, *then*, and *meanwhile* (Liu, 2008, pp. 512–517).
- (4) The multiword expressions *a bit* and *a little* were excluded when they occurred as modifier collocates, as they function as fixed phrases rather than single-word modifiers.
- (5) Multifunctional items were evaluated in context, as some forms function either as independent adverbs or as particles in phrasal verbs (e.g., *forward*). When such items formed part of a recognized phrasal verb, this was verified against the *Oxford Learner's Dictionaries* (Oxford University Press, n.d.), and the item was excluded from the modifier category.

The methodological decision to exclude time, frequency, and place adverbs was motivated by the distinction between complements and adjuncts in valency theory. Adjuncts are characterized by the fact that “they can occur relatively freely and they are not determined in their form by the governing verb” (Herbst, 2004, p. xxiv). Time, frequency, and place expressions typically exhibit these characteristics and were therefore excluded from the modifier category.

A third principle concerned object collocates. Nouns were excluded when they functioned primarily as units of measurement, currency, quantity, or numerical expressions (e.g., *pound*, *quid*, and *percent*), as well as when they formed part of fixed expressions (e.g., *a couple/lot/bit of*). Collocates referring to people, institutions, or places—especially in NEWS and PARL—were also excluded if they occurred mainly as fragments of honorific titles, personal names, geographic names, or abbreviations. In contrast, nouns denoting social roles, offices, groups, institutions, documents, or parliamentary entities were retained when they functioned independently as lexical categories, such as *Secretary*, *Chancellor*, *House*, and *Treasury*.

A fourth principle concerned the validation of complement collocates. For lexical verbs that can be followed by adjectival complements (such as *get*), the Word Sketch category “*Adjectives after X*” was additionally checked. When the adjective functioned as a complement of the node verb, it was retained and classified as a complement collocate.

Finally, transcription artefacts or non-words, such as *s*, as well as cases where the node verb appeared as a self-collocate were excluded. Orthographic and inflectional variants were grouped under a shared lemma, such as *Minister*, *minister*, and *Ministers*, with raw frequencies summed across variants and the highest logDice value used as the representative score.

3.5 Analytical procedures

In the study, **grammatical relations** refer to recurrent syntagmatic relations identified through Word Sketch, such as *objects of X*. **Grammatical behaviour** refers to the range of major grammatical relations in which a node verb occurs in the Word Sketch output. **Relational collocations** refer to individual lexical collocates occurring within these grammatical structures for example *word* as the object collocate of *say*. **Collocational profiles** refer to the aggregate set of statistically filtered and manually validated relational collocations associated with a given lexical verb across its attested grammatical relations. For instance, a collocational profile for *say* includes object collocates such as *word* and *sentence* and modifier collocates such as *rightly*. The broader term **collocational pattern** is used to refer to distributional tendencies observed across grammatical relations, relational collocations, and collocational profiles. On the basis of these definitions, the analysis of verb collocations across registers was organised around the three research questions.

3.5.1 Analysis of register variation in verb collocations

To address the first research question, which concerns the extent to which verb collocations vary across registers, the grammatical behaviour of the 25 lexical verbs was first compared across the four registers. This analysis focused on the number of major grammatical relations represented in the Word Sketch output, with raw frequencies documented. Although Word Sketch also provides information on some usage patterns (e.g. passive uses), these were excluded from the main analysis. Preliminary validation of passive uses of *say*, reported in **Section 4.3.2**, showed a high proportion of false positives, suggesting that these automatically generated categories were not reliable enough for the present collocational analysis.

A case study of *say* was then conducted as an initial in-depth analysis of register variation in a high-frequency lexical verb, before the analysis was extended to the full set of target verbs. Based on the grammatical behaviour of *say*, six grammatical relations were selected for detailed analysis because they were frequent and sufficiently comparable across registers: *X and/or complements of X*, *infinitive objects of X*, *modifiers of X*, *objects of X*, and *subjects of X*. In this section, the tagged data were evaluated, including collocates detected within these six grammatical relations as well as passive uses of *say*. Using the statistical thresholds and validation procedures described in **Section 3.4**, a validated dataset of relational collocations was compiled. This dataset was then used to examine how the collocational patterns of *say* varied across registers.

One methodological limitation should be noted. Owing to the retrieval limits of downloadable Word Sketch outputs, the collocate inventory was restricted to a maximum of 100 collocates per grammatical relation, ranked by logDice. As grammatical-relation frequencies were derived from these outputs, this ceiling may have affected frequency counts for highly productive relations. However, since the same limit was applied consistently across verbs and registers, cross-register comparability was largely maintained.

3.5.2 Analysis of frequency, distribution, and register sharing

To address the second research question, which examines how register influences the frequency and distribution of verb collocations, the analysis was extended to the validated relational collocations of the 25 lexical verbs. This stage focused on three core grammatical relations: *objects of X*, *modifiers of X*, and *complements of X*.

The analysis considered both raw frequencies and proportional distributions, with the latter calculated by normalising the raw frequency of relational collocations against the total number of verb occurrences in each register. In addition, relational collocations were classified according to their degree of register sharing: collocations shared across four registers, collocations shared across three registers, collocations shared across two registers, and collocations specific to one register. This classification enabled the identification of both broadly shared collocational patterns and those that are distinctive to particular registers.

3.5.3 Semantic and functional interpretation of collocational patterns

To address the third research question, which concerns how collocational patterns relate to the communicative purposes and discourse functions of individual registers, object collocates of the 25 verbs were subjected to semantic analysis. This approach was methodologically motivated by Han and Shin's (2015) register-based semantic analysis of verb collocates and conceptually informed by Corpus Pattern Analysis (CPA), as discussed in **Section 2.4.2** and **Section 2.2.4.3**, respectively.

Semantic analysis was used to identify the semantic domains represented by the object collocates and to compare their distribution across registers. Register-specific concentrations of particular semantic domains were then used as the basis for functional interpretation, linking semantic preferences to the communicative purposes and discourse functions of CONV, PARL, NEWS, and ACAD. Furthermore, shared object collocates were analysed according to their degree of register sharing, that is, whether they occurred across four, three, or two registers, while register-specific object collocates were compared across the four registers. Representative relational collocations and selected concordance lines were further examined to illustrate how these semantic patterns were realised in context.

Semantic annotation was carried out using the USAS online English tagger (<https://ucrel-api.lancaster.ac.uk/usas/tagger.html>), which implements the UCREL Semantic Analysis System (USAS). USAS provides an automated framework for semantic annotation (Rayson, 2008). Its original tagset was based on McArthur's (1981) *Longman Lexicon of Contemporary English*, while the revised tagset is hierarchically organised into 21 major discourse fields, as shown in Figure 3. These major discourse fields are further divided into 232 category labels, as listed on the website (<https://ucrel.lancs.ac.uk/usas/>). For example, the discourse field **T: Time** includes four semantic subcategories, such as **T1: Time**, **T2: Time: Beginning and ending**, **T3: Time: Old, new and young; age**, and **T4: Time: Early/late**. The subcategory **T1: Time** is further subdivided into six more specific labels, including **T1.1: Time: General**, **T1.2: Time: Momentary**, and **T1.3: Time: Period**, as illustrated in Figure 4.

A general and abstract terms	B the body and the individual	C arts and crafts	E emotion
F food and farming	G government and public	H architecture, housing and the home	I money and commerce in industry
K entertainment, sports and games	L life and living things	M movement, location, travel and transport	N numbers and measurement
O substances, materials, objects and equipment	P education	Q language and communication	S social actions, states and processes
T Time	W world and environment	X psychological actions, states and processes	Y science and technology
Z names and grammar			

Figure 3: 21 discourse fields in USAS tagset

T TIME

T1	Time
T1.1	Time: General
T1.1.1	Time: General: Past
T1.1.2	Time: General: Present; simultaneous
T1.1.3	Time: General: Future
T1.2	Time: Momentary
T1.3	Time: Period
T2	Time: Beginning and ending
T3	Time: Old, new and young; age
T4	Time: Early/late

Figure 4: Semantic subcategories of T: Time

In the USAS output, each item is assigned to one syntactic tag and one or more semantic tags. When more than one tag is possible, the tags are presented in approximate order of likelihood, with the most probable tag appearing first. Some tags also include additional symbols, such as plus and minus signs (i.e., +, ++, +++; -, --, ---), which indicate a positive or negative position on a semantic scale, such as A7 (Probability), A7+ (Likely), A7- (Unlikely) (Rayson et al., 2002).

For the present analysis, the first USAS tag assigned to each collocate was used as the basis for semantic annotation, as this represents the most probable interpretation in the USAS output. This principle was also applied when alternative tags were provided, such as Q2.1/S2, in which

case only the first tag was retained. Each retained tag was then mapped onto two levels of classification adopted in the present study: a more specific semantic subdomain and a broader semantic domain. For example, as shown in Table 13, the tag Q3 is recorded as the semantic subdomain Q3: Language, speech and grammar, and is grouped under the broader semantic domain Q: Language and communication. Thus, the annotation scheme recorded three types of information for each collocate: the original USAS semantic tag, the semantic domain, and the semantic subdomain. The analysis across the 25 verbs focused primarily on the broader domains to facilitate generalization, while semantic subdomains were used where more detailed interpretation was required, especially in the case study of *say*.

Table 13

Semantic Annotation of Illustrative Examples

Grammatical relation	Relational collocation	Collocate	USAS semantic tag	Semantic domain	Semantic subdomain
Subjects of X	<i>speaker + say</i>	speaker	Q2.1/S2mf G1.1/S2mf O3	Q: Language and communication	Q2: Speech acts
Objects of X	<i>say + word</i>	word	Q3 Q2.1 S9@	Q: Language and communication	Q3: Language, speech and grammar
Modifiers of X	<i>explicitly + say</i>	explicitly	A4.2+	A: General and abstract terms	A4: Classification
X and/or ...	<i>write and say</i>	write	Q1.2 I2.1	Q: Language and communication	Q1: Communication
Infinitive objects of X	<i>say + represent</i>	represent	Q1.1 A6.2+ S1.1.3 A5.4- K4	Q: Language and communication	Q1: Communication
Complements of X	<i>make + clear</i>	clear	A7+ A10+ O4.3 O4.1 N5.1+ W4	A: General and abstract terms	A7: Definite

3.5.4 Descriptive and inferential statistics

Descriptive statistics were first used to summarise the distribution of grammatical relations across registers. Given that the four corpora differed in size and the target verbs themselves occurred with different frequencies across corpora, cross-register comparisons were based primarily on relative frequencies. For each target verb, the relative frequency of a grammatical relation was calculated as the number of occurrences of that relation divided by the total number of occurrences of the verb in the corresponding register and expressed as a percentage. Raw frequencies were also reported where relevant.

Descriptive statistics were also used to summarise the distribution of relational collocations. At this level, both token counts and type counts were reported, depending on the analytical focus. Type counts were used to indicate the number of distinct relational collocates attested in a given register, grammatical relation, or degree of register sharing, while percentages were used to show their proportional distribution across the relevant categories.

To determine whether the observed register differences in grammatical relations were statistically significant, inferential analyses were subsequently conducted for each verb. For each target verb, the denominator used in the statistical comparisons was the total number of occurrences of that verb in the respective register, rather than the total number of words in the register. This denominator was adopted because the analysis was concerned with variation in the grammatical behaviour of individual verbs across registers rather than differences in their overall

corpus frequency. Pairwise comparisons were then conducted across all register combinations, such as CONV versus PARL.

Log-likelihood tests were used to assess the statistical significance of each pairwise difference because they are a standard method for comparing frequency distributions across corpora of different sizes and do not assume a normal distribution (McEnery & Hardie, 2011). Log Ratio was used to measure the magnitude and direction of the difference (Hardie, 2014). A Log Ratio of 1 corresponds to a twofold difference, while a value of 2 corresponds to a fourfold difference. Positive values indicate overuse in the first register in a pairwise comparison, while negative values indicate overuse in the second register.

As the analysis was pairwise, overuse was defined relative to a specific comparison register rather than to the other three registers collectively. For example, a relation could be overused in CONV relative to NEWS but not relative to PARL or ACAD. This pairwise comparison design yielded six register comparisons for each grammatical relation. To account for the resulting multiple testing problem, a Bonferroni correction was applied. With an overall significance level of $\alpha = .05$, the corrected significance threshold was $p < .0083$. Pairwise differences were therefore retained as cases of overuse only when they met this adjusted significance threshold and had an absolute Log Ratio greater than 1 ($|\text{Log Ratio}| > 1$). Relations that met these criteria for one register relative to each of the other three registers were subsequently classified as register-distinctive patterns.

3.6 Conclusion

This chapter has described the methodology of the study, covering corpora selection, the identification of target lexical verbs, the extraction and validation of relational collocations, and the analytical procedures used to examine their distribution, register variation, and semantic interpretation. The next chapter presents the results of these analyses.

4 Results

4.1 Introduction

The chapter presents the results of the grammatical and collocational analysis of the 25 lexical verbs across registers. The analysis proceeds from a broad overview of grammatical behaviour to a more detailed examination of relational collocations, with each stage addressing different research questions. **Section 4.2** first provides an overview of the grammatical behaviour of the 25 lexical verbs based on automatically extracted data, including the distribution of attested grammatical relations and register-distinctive patterns.

Given that the accuracy of the automatically extracted data affects the reliability of the identified patterns, **Section 4.3** presents a case study of *say*, the most frequent verb in the dataset. The case study first describes the distribution of the attested grammatical relations of *say* across registers on the basis of the automatically extracted data. It then evaluates the tagged data and manually validates the relational collocations occurring in the six selected relations. The analysis subsequently examines the distribution of the validated relational collocations across grammatical relations and registers, followed by analyses of shared and register-specific collocations and their semantic patterns.

Sections 4.4 and **4.5** extend the analysis of relational collocations to all 25 lexical verbs. Section 4.4 examines the distribution of relational collocations across the four registers and three selected grammatical relations. Section 4.5 then investigates the extent to which these collocations are shared across registers or specific to individual registers.

The scope of the grammatical analysis therefore becomes progressively more focused across these stages. Section 4.2 considers the major grammatical relations attested for the 25 lexical verbs, while the case study in Section 4.3 examines six selected relations in greater detail. The cross-verb analyses in Sections 4.4 and 4.5 concentrate on three relations, with the semantic analysis in **Section 4.5.6** restricted to relational collocations in the object relation.

Supplementary results are provided in the appendices. **Appendices C** and **D** report the statistical results for the major grammatical relations and the three relations, respectively. **Appendix E** provides the full lists of relational collocations, including shared and register-specific collocations for the 25 lexical verbs, together with the semantic classifications used in the analysis.

4.2 Grammatical Relations across the 25 Verbs: Distribution and Register-Distinctive Patterns

4.2.1 Distribution of attested grammatical relations

This section provides an overview of the number of grammatical relations in which each lexical verb occurs across the four registers. For ease of presentation and comparison, all attested grammatical relations were grouped into two broad types in Table 14, namely, general clause-level grammatical relations and particle-related grammatical relations.

The first category covers broadly comparable clause-level patterns, including subjects, objects, modifiers, coordination relations, infinitive objects, complements, prepositional relations, *wh*-patterns, pronominal subjects, pronominal objects, and *-ing* objects. Apart from the first six relations reported in Table 12, other relations are illustrated in (1) below.

(1)	General clause-level grammatical relations
a.	Liege said in a statement on their website that Fryers "felt homesick and had difficulty acclimatising to life away from home and family. (<i>prepositional relations</i>)
b.	Cassells says what the school teaches goes no further than "daily and seasonal routines, and diet, for making you more balanced. (<i>wh-patterns</i>)
c.	He said the missile made a crater a metre deep. (<i>pronominal subjects</i>)
d.	Says one: I've lived here for 30 years. (<i>pronominal objects</i>)
e.	And anyway, who says cooking frozen chips or testing the limits of a shatterproof ruler isn't action? (<i>-ing objects</i>)

The second category comprises particle-related grammatical relations, including Word Sketch relations involving a particle following the verb and relations labelled as objects of verb–particle combinations, as illustrated in (2). In addition, the counts are based on Word Sketch relation headings rather than on the number of individual particle collocates listed under each heading.

(2)	Particles and particle-verb patterns
a.	They will say it over and over again. (<i>particles after "say" with object</i>)
b.	The White House said around 6,000 Americans have been evacuated from Afghanistan since 14 August. (<i>objects of "say around"</i>)
c.	"Toast," he says down the satellite phone . (<i>objects of "say down"</i>)
d.	But one senior Conservative said off the record : "The war hadn't featured high up on the list of people's priorities, but by the end it was the third most important. (<i>objects of "say off"</i>)

As shown in Table 14, the overall number of attested grammatical relations varies across registers. NEWS records the highest totals in both categories, with 300 general clause-level grammatical relations and 136 particle-related grammatical relations. PARL follows, with 293 and 102 respectively. Lower totals are found in ACAD and CONV. ACAD has 270 general clause-level grammatical relations and 52 particle-related grammatical relations, while CONV has 265 and 64 respectively. The cross-register differences are therefore relatively limited for general clause-level relations, which range from 265 to 300, but more pronounced for particle-related grammatical relations, which range from 52 to 136.

A similar pattern can be observed at the verb level. NEWS shows the broadest grammatical coverage, with all 25 verbs attested across the full set of 12 general clause-level grammatical relations. PARL displays a comparable pattern, as most verbs occur in all or nearly all of these relations. By contrast, CONV and ACAD show more limited coverage. In both registers, fewer than half of the verbs reach the maximum of 12 general clause-level relations. This reduced coverage is particularly visible for verbs such as *bring*, *show*, *call*, and *ask* in CONV, and for verbs such as *call*, *go*, and *look* in ACAD.

Overall, NEWS and PARL show slightly broader grammatical coverage, whereas CONV and ACAD are more restricted. Cross-register differences are relatively small for general clause-level relations but more evident in particle-related relations. At the verb level, the extent of grammatical coverage also varies across registers, suggesting that individual lexical verbs show preferences for different grammatical relations in different registers. Building on this distributional overview, the following section examines which grammatical relations are significantly overused for particular verbs in one register relative to the other three.

Table 14

Relative Frequencies and Number of Grammatical Relations Attested per Verb across Registers

Verb	Measure	CONV	PARL	NEWS	ACAD
say	RF	4,714.47	2,704.14	4,552.80	780.23
	GCR / PGR	12 / 1	12 / 3	12 / 3	11 / 0
make	RF	1,542.31	3,156.26	1,921.26	1,735.43
	GCR / PGR	10 / 3	12 / 3	12 / 6	12 / 3
take	RF	1,470.22	2,000.13	1,586.89	1,175.33
	GCR / PGR	12 / 4	12 / 8	12 / 9	11 / 4
see	RF	2,450.03	1,196.52	1,135.00	1,283.30
	GCR / PGR	12 / 2	12 / 4	12 / 5	11 / 0
go	RF	6,961.17	1,621.24	1,738.68	509.94
	GCR / PGR	11 / 5	12 / 8	12 / 10	9 / 2
know	RF	7,793.17	1,279.24	929.21	619.78
	GCR / PGR	12 / 2	11 / 0	12 / 2	10 / 0
give	RF	822.70	1,585.53	977.91	1,253.46
	GCR / PGR	9 / 3	12 / 5	12 / 6	12 / 3
come	RF	2,119.25	1,199.98	1,183.82	528.89
	GCR / PGR	11 / 5	11 / 6	12 / 7	10 / 2
think	RF	5,977.55	1,137.62	901.99	414.40
	GCR / PGR	12 / 2	12 / 2	12 / 3	11 / 2
use	RF	1,300.52	704.42	724.61	1,407.67
	GCR / PGR	9 / 2	12 / 3	12 / 3	12 / 2
get	RF	8,591.78	1,314.10	1,673.84	290.94
	GCR / PGR	12 / 5	12 / 7	12 / 9	12 / 3
need	RF	1,264.27	1,629.18	677.57	471.08
	GCR / PGR	12 / 0	12 / 2	12 / 2	12 / 0
work	RF	904.26	1,189.22	669.88	437.43
	GCR / PGR	10 / 3	11 / 5	12 / 8	11 / 3
find	RF	671.09	534.03	816.11	864.65
	GCR / PGR	12 / 3	12 / 3	12 / 5	12 / 3
look	RF	1,680.99	962.71	787.90	296.78
	GCR / PGR	12 / 3	11 / 6	12 / 9	9 / 3
leave	RF	564.95	676.62	672.72	313.24
	GCR / PGR	10 / 2	12 / 6	12 / 7	12 / 4
put	RF	1,434.05	805.55	541.65	263.08
	GCR / PGR	11 / 2	12 / 8	12 / 11	10 / 3

Continued on next page

Table 14 continued

Verb	Measure	CONV	PARL	NEWS	ACAD
<i>mean</i>	RF	2,191.26	458.97	371.78	412.98
	GCR / PGR	11 / 0	12 / 1	12 / 2	11 / 0
<i>want</i>	RF	2,050.29	1,242.28	830.64	201.69
	GCR / PGR	10 / 2	11 / 2	12 / 2	10 / 1
<i>pay</i>	RF	556.16	544.35	435.61	281.01
	GCR / PGR	12 / 4	12 / 6	12 / 7	11 / 3
<i>call</i>	RF	618.61	329.68	507.91	320.16
	GCR / PGR	9 / 2	12 / 4	12 / 5	7 / 2
<i>ask</i>	RF	440.80	750.08	454.99	245.38
	GCR / PGR	9 / 1	11 / 1	12 / 2	10 / 0
<i>show</i>	RF	223.28	335.89	515.09	864.71
	GCR / PGR	7 / 2	11 / 2	12 / 4	12 / 2
<i>try</i>	RF	826.42	442.78	496.43	222.57
	GCR / PGR	11 / 2	12 / 2	12 / 3	11 / 3
<i>bring</i>	RF	269.08	618.39	383.95	285.89
	GCR / PGR	7 / 4	12 / 5	12 / 6	11 / 4
Total	GCR / PGR	265 / 64	293 / 102	300 / 136	270 / 52

Note. RF = relative frequency per million words; GCR = general clause-level grammatical relations; PGR = particle-related grammatical relations.

4.2.2 Register-distinctive patterns across the 25 verbs

As described in **Section 3.5.4**, significant overuse refers to a grammatical relation attested for a given verb that occurs significantly more frequently in one register than in a specific comparison register, while a register-distinctive pattern refers to a relation associated with a verb that is significantly overused in one register relative to each of the other three registers. The full pairwise results are provided in **Appendix C**. For reasons of space, Table 15 reports only the strongest cross-register contrasts in which the same grammatical relation is overused in one register relative to all three others.

As shown in Table 15, register-distinctive grammatical relations vary considerably across both registers and lexical verbs. At least one register-distinctive relation was identified for 21 of the 25 verbs in CONV, compared with 17 in ACAD, 15 in NEWS, and only five in PARL.

In CONV, *pronominal objects of X* are register-distinctive for 17 verbs and *pronominal subjects of X* for 10 verbs, while *wh-words following X* are register-distinctive for four verbs. The first two relations indicate a strong tendency for register-distinctive overuse involving pronominal arguments. Notably, CONV is the only register in which *wh-words following X* are register-distinctively overused, for *call*, *leave*, *look*, and *mean*. PARL, by contrast, shows considerably fewer register-distinctive relations, as only five verbs are involved and no individual relation is distinctive for more than two verbs.

NEWS displays a different distribution, with register-distinctive overuse concentrated particularly in particle-related constructions. *Objects of “X off”* are register-distinctive for 10 verbs, while *objects of “X out”* and *“X up”* are each distinctive for five verbs. In ACAD, *infinitive objects of X* constitute the most frequent register-distinctive relation, occurring for 11 verbs. Alongside this pattern, ACAD also shows considerable verb-specific variation, with several other

register-distinctive relations occurring for only one verb.

Register effects are also evident at the level of individual lexical verbs. In particular, *leave*, *put*, and *work* each show at least one grammatical relation that is significantly overused in a given register relative to all three other registers, although the specific relations differ by register. For example, for *work*, pronominal objects, pronominal subjects, and *X and/or* are register-distinctively overused in CONV; infinitive objects in PARL; several particle-related object relations and complements in NEWS; and *objects of "X out"* and *objects of X* in ACAD.

Overall, these results provide preliminary evidence that the grammatical behaviour of the 25 lexical verbs varies systematically across registers. The clearest register-level contrasts involve pronominal and interrogative relations in CONV, particle-related relations in NEWS, and infinitive objects in ACAD, whereas PARL shows fewer register-distinctive contrasts. Meanwhile, variation at the level of individual verbs suggests that the same verb may exhibit different register-specific grammatical preferences, reflected in the significant overuse of different relations across registers.

While these results provide an initial overview of register variation in the grammatical behaviour of the 25 verbs, they are based on automatically extracted Word Sketch relations and thus require further validation. To address this limitation, Section 4.3 turns to a case study of *say*, providing a more detailed analysis based on manually validated relational collocations.

Table 15

Register-Distinctive Grammatical Relations across the 25 Verbs

Register	Overused grammatical relation	<i>n</i>	Verbs
CONV	Pronominal objects of X	17	ask, bring, call, come, give, leave , look, mean, need, pay, say, show, take, try, use, want, work
	Pronominal subjects of X	10	bring, give, make, mean, put , say, show, take, use, work
	X and/or	4	come, find, try, work
	Wh-words following X	4	call, leave , look, mean
	Objects of "X out"	3	come, find, look
	-ing objects of X	2	know, look
	Infinitive objects of X	1	use
	Objects of "X up"	1	come
	Particles after X	1	find
PARL	Particles after X with object	1	say
	Objects of "X out"	2	call, leave
	X and/or	1	put
	Infinitive objects of X	1	work
NEWS	Objects of "X around"	1	get
	Objects of "X off"	10	call, come, give, go, pay, put , see, show, take, work
	Objects of "X out"	5	give, go, pay, put , see
	Objects of "X up"	5	call, give, show, think, work
	-ing objects of X	3	leave , see, want
	Complements of X	3	come, try, work
	Particles after X with object	3	go, need, see
	Objects of "X away"	2	give, put
ACAD	Objects of "X over"	1	take

Continued on next page

Table 15 continued

Register	Overused grammatical relation	<i>n</i>	Verbs
	Infinitive objects of X	11	bring, come, find, go, know, make, put , say, see, show, think
	X and/or	5	go, know, say, think, want
	Prepositional phrases	4	find, go, need, show
	Adjectives after X	2	go, think
	-ing objects of X	1	need
	Complements of X	1	think
	Modifiers of X	1	go
	Objects of “X aside”	1	leave
	Objects of “X open”	1	leave
	Objects of “X out”	1	work
	Objects of “X up”	1	look
	Objects of X	1	work
	Particles after X	1	try
	Particles after X with object	1	think
	Subjects of X	1	go

Note. *n* = number of verbs. Overused grammatical relations were identified using a Bonferroni-adjusted significance threshold of $p < .0083$ and an absolute Log Ratio greater than 1 ($|\text{Log Ratio}| > 1$).

4.3 Case Study: Grammatical and Collocational Patterns of *Say*

This section presents a case study of *say* as an illustrative example of the analytical procedures applied to the wider set of lexical verbs. The analysis moves from its grammatical behaviour and evaluation of the tagged data to the distribution of manually validated relational collocations across six selected grammatical relations, followed by analyses of shared and register-specific collocations and an interim summary.

4.3.1 Grammatical behaviour of *say* across registers

Table 16 presents the distribution of 18 attested grammatical relations of *say* across the four registers, with percentages calculated relative to all grammatical-relations tokens of *say* within each register. In CONV, *say* most frequently occurs with pronominal subjects (48.58%), followed by modifiers (17.41%) and objects (10.45%). In PARL, subjects (29.10%) and pronominal subjects (29.03%) are similarly frequent, followed by modifiers (13.82%) and prepositional phrases (13.77%). In the written registers, broadly comparable patterns are observed, with pronominal subjects, subjects, and objects among the most frequent relations. However, NEWS shows a lower proportion of modifiers (4.19%), whereas ACAD displays comparatively higher proportions of modifiers and prepositional phrases, resembling PARL. Infinitive objects are also more frequent in ACAD than in the other registers.

Table 17 shows that the reflexive and passive uses of *say* occur primarily in NEWS, and to a lesser extent in PARL, whereas they are rare or absent in CONV and ACAD. Relevant examples are shown in Example (3) below. Given the unexpectedly limited occurrence of passive uses in ACAD, these instances are additionally evaluated in the next section.

Overall, these results suggest that the grammatical realization of *say* is shaped by register, but the contrasts are relation-specific rather than uniform across all pairwise comparisons.

(3)	Reflexive and passive uses of <i>say</i>
a.	It is something I want to say myself. (<i>As reflexive</i>)
b.	There was nothing said of any other problem. (<i>In passive</i>)

Table 16
Distribution of Major Grammatical Relations for say across Registers

Grammatical relation	CONV (%)	PARL (%)	NEWS (%)	ACAD (%)
Modifiers of X	17.41	13.82	4.19	12.19
Objects of X	10.45	8.49	24.76	19.69
Subjects of X	5.94	29.10	36.39	21.51
X and/or	0.29	0.42	0.36	1.65
Prepositional phrases	2.26	13.77	3.29	9.96
Particles after X	—	0.03	0.02	—
Particles after X with object	0.04	—	0.01	—
Pronominal objects of X	9.93	1.75	3.95	—
Pronominal subjects of X	48.58	29.03	24.54	21.22
Wh-words following X	3.63	2.62	0.62	2.71
Infinitive objects of X	0.28	0.28	1.20	9.75
-ing objects of X	0.36	0.06	0.30	—
Complements of X	0.18	0.19	0.13	0.52
Adjectives after X	0.65	0.42	0.24	0.80
Objects of “X up”	—	< 0.01	—	—
Objects of “X off”	—	—	< 0.01	—
Objects of “X around”	—	—	< 0.01	—
Objects of “X down”	—	—	< 0.01	—

Note. Percentages represent the proportion of each grammatical relation relative to all grammatical-relation tokens within a register. A dash indicates that no tokens were observed.

Table 17
Usage Patterns Identified for say across Registers

		Raw frequency			
Verb	Usage pattern	CONV	PARL	NEWS	ACAD
<i>say</i>	As reflexive	—	66	340	—
	In passive	7	17	242	—

Note. — indicates no occurrences recorded.

4.3.2 Evaluation of tagged data

This section evaluates the automatically extracted Word Sketch data used in the subsequent analysis. The evaluation was conducted at the level of candidate relational collocation entries, each representing a collocate identified under a specific grammatical relation for *say*, together with its frequency and logDice score. The entries were manually checked and classified as retained, merged, or excluded according to the validation criteria described in **Section 3.4**. The results are presented as the number and percentage of excluded entries in each register. The full validation record is provided in **Appendix B**.

Figure 5 presents the concordance lines for *say* + *point*, where *point* was automatically identified as an object collocate of *say*. However, manual inspection of the ten concordances shows that *point* does not generally function as an independent object of *say*. Six instances, namely lines 3, 4, 6, 7, 8, and 9, occur in numerical expressions such as *said one point five*, where *point* functions as a decimal marker within a larger number expression. Line 10, *say fifty points wild*, appears to refer to a score or quantity. In contrast, Lines 2 and 5 contain the idiomatic expression *no point*, where *point* is the head of an object structure. Line 1, *he said the point that*, is potentially close to an object use. Since these three cases are not representative of the overall pattern, *point* was excluded from the set of object collocates of *say*. This example shows that some automatically extracted collocates are valid lexical items but are not valid relational collocations in the grammatical relations assigned by the parser.

	Left context	KWIC	Right context
1	Discussing Expl... ough er erm you know so everything he says I don't take it seriously but erm like I	said the .point that he is he does he does share the family surname so er yeah there	
2	Discussing Expl... ving a starter is he not? no I said <unclea> oh I said don't you want a starter? he	says no .point being greedy I'm gonna have a nice big steak and a pudding no point t	
3	Discussing Expl... uted to the fracking mm so that's g now then there's this one in Lincolnshire and I	say three .point four for this country is is er getting on the big side for us and of coun	
4	Discussing Expl... int one by two dunno what the fuck that means is that is that centimetres? yeah it	says two .point five inch though it's gotta be bigger than that cos that's fucking tiny wf	
5	Discussing Expl... because I said to what are you gonna do for forty-five minutes after school now? I	said no .point coming home cos as soon as you walk through the door we'd have to t	
6	Discussing Expl... m was the closest what was it one point five? it was one point four yeah I thought	said one .point five yeah but he changed his bet halfway though oh right I didn't she's	
7	Discussing Expl... roes did it cost? erm I don't know about resources all I know is the figure that they	said the one .point seven million that's a lot of money for one individual where erm wi	
8	Discussing Expl... ave er have a look unless she hasn't because we're still going straight that way it	says one .point eight miles until you change oh right well just keep your eye on it she'	
9	Discussing Expl... hu won't we in the next week? probably I don't know how much it is here I think it	said eight .point four so less than a pound not a fan of that man's Mohican no he hea	
10	Discussing Expl... oo many no cos there are extra ones there's ones there's jokers and there's one it	says fifty .points wild <unclea> <unclea> it's your dog good	

Figure 5: Concordance lines for *say* + *point* in the BNC2014 spoken corpus

The case of *say* + *point* illustrates one type of mismatch identified during validation. The same manual checking procedure was applied to all candidate relational collocations, and the overall results are summarized in Table 18. In total, 1,183 candidate entries were checked across the four registers. Of these, 449 were retained as lemma entries, 10 were merged as orthographic variants, and 724 were excluded according to the validation criteria.

The relatively high exclusion rate reflects both the deliberately restrictive validation procedure adopted in this study and errors in automatic tagging or relation assignment. A substantial proportion of the exclusions resulted from the validation procedure itself, under which candidate entries were removed when they did not meet the criteria, including time, place, or frequency adverbs.

In addition, genuine parsing and relation-assignment errors were identified during manual checking. These errors were unevenly distributed across grammatical relations and registers, with objects and complements showing more mismatches than some other relations and parsing problems being particularly noticeable in CONV. The proportion of excluded entries also varied across registers, from 53.66% in ACAD, 54.48% in PARL, 63.01% in NEWS, to 68.35% in CONV. Merged variant entries occurred only in PARL and NEWS. Notably, all candidate collocates identified under the relation *complements of X* were excluded because they did not function

as complements directly dependent on the verb *say*. In Example (3a), (3b) and (3d), *different*, *positive*, and *new* occur within the object noun phrase headed by *something* and modify that pronoun rather than the verb *say*. They therefore function as postmodifiers within the noun phrase, not as complements of *say*. Example (3c) represents a different mismatch: *international* belongs to the reporting noun phrase *the Norway international* following *said* and is not syntactically related to *say* as its complement.

(3)	Examples of excluded cases of complement
a.	and then the next day you were saying something completely different without discussing anything again (CONV)
b.	I hope he will also say something positive about enhancing animal welfare standards. (PARL)
c.	"I knew before the season started it was going to be a challenge for me this year," said the Norway international . (NEWS)
d.	An article is more likely to be accepted if it says something new but this might annoy some of his party leaders. (ACAD)

The validation results reveal the limitations in the automatic extraction of relational collocations in Sketch Engine, with some candidate entries failing to correspond to the grammatical relations assigned in the Word Sketch output. These mismatches were more frequent in certain relations, particularly objects and complements, and may also reflect the syntactic flexibility of the high-frequency verb *say*. The relatively high exclusion rate therefore highlights the importance of manual validation in establishing a reliable dataset for register-based collocational analysis.

Table 18

Excluded Entries in Selected Automatically Extracted Grammatical Relations for Say

Register	Candidate entries checked	Retained lemma entries	Merged variant entries	Excluded entries	Excluded entries (%)
CONV	278	88	0	190	68.35
PARL	290	126	6	158	54.48
NEWS	492	178	4	310	63.01
ACAD	123	57	0	66	53.66
Total	1,183	449	10	724	61.20

Note. Percentages were calculated by dividing the number of excluded entries by the number of candidate entries checked within each register.

A further evaluation was conducted on automatically identified passive uses of *say* to assess the reliability of broader usage-pattern categories automatically generated by the parser, which also required validation before being considered for analysis. Table 19 shows that 266 entries were identified as passive uses, but manual validation retained only 36 cases and excluded 230, resulting in an overall exclusion rate of 86.47%. No passive entries were found in ACAD; all seven CONV entries were excluded, while 4 of the 17 PARL entries and 32 out of the 242 NEWS entries were retained. The full validation record is also provided in **Appendix B1**.

The results indicate that the automatic identification of passive uses of *say* was not reliable enough to support register-based analysis, as most retrieved items were tagging or parsing errors rather than genuine passive constructions. Passive-use results are therefore treated only as a

supplementary validation check, supporting the exclusion of usage-pattern categories from the subsequent quantitative analysis. The main analysis instead focused on manually validated relational collocations.

Table 19

Validation of Automatically Identified Passive Uses of Say by Register

Register	Automatically identified passive entries	Retained passive entries	Excluded passive entries	Excluded passive entries (%)
CONV	7	0	7	100.00%
PARL	17	4	13	76.%
NEWS	242	32	210	86.78%
ACAD	0	0	0	—
Total	266	36	230	86.47%

Note. Percentages were calculated by dividing the number of excluded passive entries by the number of automatically identified passive entries within each register. A dash indicates that the percentage was not calculated because no passive entries were automatically identified.

4.3.3 Distribution of relational collocations for *say* across registers and six selected grammatical relations

Six grammatical relations—*X and/or, complements of X, infinitive objects of X, modifiers of X, objects of X, subjects of X*—were selected to examine the distribution of relational collocations for *say* across the four registers. These relations were chosen because they represent general clause-level grammatical relations and contain lexical collocates suitable for further analysis.

Table 20 summarises the overall distribution of validated relational collocations of *say* across registers. The number of distinct collocates varies considerably, with NEWS recording the largest number (178), followed by PARL (126) and CONV (88), and ACAD (57). NEWS therefore shows the greatest variety of relational collocations, partly reflecting the high overall frequency of *say* in this register. However, given that the overall frequency of *say* varies across registers, percentages relative to the total occurrences of *say* in each register are used for cross-register comparison. On this measure, relational collocations account for the largest share of *say* in PARL (15.52%), followed by NEWS (9.85%), ACAD (7.26%), and CONV (5.38%). This suggests that relational collocations constitute a small proportion of all occurrences of *say* in each register, but are most prominent in PARL.

Table 20

Overall Distribution of Relational Collocations for Say by Register

Register	Total no. of collocates	% of <i>say</i>
CONV	88	5.38
PARL	126	15.52
NEWS	178	9.85
ACAD	57	7.26

Note. Percentages were calculated using the total number of occurrences of *say* in each register as the denominator.

Following the overall cross-register distribution presented in Table 20, Table 21 breaks down

the validated relational collocations of *say* by grammatical relation within each register, using the number of distinct collocates and their percentages of all occurrences of *say* as the measures for comparison.

Across the six relations, *subjects of X* account for the largest proportions of *say* in PARL (11.01%) and NEWS (8.14%), while their proportions are lower in CONV (3.18%) and ACAD (2.52%). *Modifiers of X* are also relatively prominent in PARL (3.20%), compared with CONV (0.92%), ACAD (0.81%), and NEWS (0.58%). By contrast, *objects of X* occur at similar proportions in CONV (1.08%) and PARL (1.07%), but are less frequent in ACAD (0.81%) and NEWS (0.39%). ACAD shows a different pattern for *infinitive objects of X*, which account for 2.52% of all occurrences of *say*, compared with 0.55% in NEWS, 0.06% in PARL, and 0.03% in CONV. *X and/or* remains infrequent in all four registers, although its proportion is comparatively higher in ACAD (0.62%). No validated relational collocations were retained for *complements of X* in any register.

Meanwhile, the number of distinct collocates does not simply follow the relative frequency of *say* across registers. Although the relative frequency of *say* decreases in the order CONV > NEWS > PARL > ACAD, *modifiers of X*, *infinitive objects of X*, and *X and/or* show comparatively greater collocational diversity in registers where *say* is not necessarily most frequent.

Table 21

Distribution of Relational Collocations of Say by Grammatical Relation and Register

Grammatical relation	Register	No. of collocates	% of <i>say</i>
X and/or	CONV	8	0.17
	PARL	25	0.17
	NEWS	64	0.19
	ACAD	3	0.62
Complements of X	–	0	0
Infinitive objects of X	CONV	1	0.03
	PARL	6	0.06
	NEWS	13	0.55
	ACAD	10	2.52
Modifiers of X	CONV	11	0.92
	PARL	40	3.20
	NEWS	37	0.58
	ACAD	10	0.81
Objects of X	CONV	12	1.08
	PARL	5	1.07
	NEWS	4	0.39
	ACAD	5	0.81
Subjects of X	CONV	56	3.18
	PARL	50	11.01
	NEWS	60	8.14
	ACAD	29	2.52

Note. No validated relational collocations were retained for *complements of X* in any register. Percentages were calculated using the total number of occurrences of *say* in each register as the denominator.

Table 22 reports the inferential results for the six selected grammatical relations, identifying cases of significant overuse in each target register relative to one or more comparison registers (full statistics are provided in **Appendix C1b**).

The grammatical behaviour of *say* varies across registers. The strongest register-distinctive contrasts occur in ACAD, where *X and/or* and *infinitive objects of X* are overused relative to all three other registers, and in PARL, where *modifiers of X* are likewise overused relative to all three. Other differences are more restricted to particular pairwise comparisons, such as *objects of X* in CONV and PARL relative to NEWS, and *subjects of X* in PARL and NEWS relative to CONV and ACAD.

Taken together, the results show that *say* varies across registers in both its collocational range and the proportion of occurrences represented by relational collocations. Its collocational range is greatest in NEWS, while relational collocations account for the largest proportion of occurrences in PARL. The inferential analysis further shows that the grammatical relations showing significant overuse differ across registers.

Table 22

Overused Grammatical Relations among the Six Selected Relations for Say across Registers

Target register	Overused grammatical relation	Overused relative to
CONV	Objects of X	NEWS
PARL	Modifiers of X	CONV; NEWS; ACAD
	Objects of X	NEWS
	Subjects of X	CONV; ACAD
NEWS	Infinitive objects of X	CONV; PARL
	Subjects of X	CONV; ACAD
ACAD	X and/or	CONV; PARL; NEWS
	Infinitive objects of X	CONV; PARL; NEWS
	Objects of X	NEWS

Note. Overused grammatical relations were identified using a Bonferroni-adjusted significance threshold of $p < .0083$ and $|\text{Log Ratio}| > 1$.

4.3.4 Shared and register-specific relational collocations of *say*

Overview

Table 23 summarizes the distribution of shared and register-specific collocates of *say* across the four registers. It includes collocate types shared across four, three, and two registers, as well as collocates unique to individual registers. For each group, the table reports the number of collocate types, a representative example, as well as the dominant grammatical relation, semantic domain, and semantic subdomain.

The results show that full cross-register sharing is limited. Only 5 collocate types occur in all four registers. In comparison, more collocates are shared across three registers (21 types) or two registers (69 types), indicating that partial overlap is more common than complete cross-register sharing. Register-specific collocates form a larger group than shared collocates, with NEWS containing the highest number of unique collocate types (97), followed by PARL (51), CONV (48), and ACAD (30). This suggests that the collocational profile of *say* in NEWS exhibits the greatest lexical diversity. Meanwhile, the dominant grammatical relations differ across collocate

groups. *X and/or* is the most frequent relation among collocates shared across four registers (40%) and NEWS-specific collocates (43.43%), while *modifiers of X* dominates those shared across three registers (42.86%). *Subjects of X* represents the largest proportion in collocates shared across two registers, and CONV-, PARL-, and ACAD-specific collocates.

In terms of semantic domains, A: General and abstract terms and Q: Language and communication are dominant in collocates shared across most categories, while CONV-specific collocates are mainly associated with S: Social actions, states and processes.

Clearer register-based distinctions emerge at the semantic subdomain level. Fully shared collocates show the highest proportion in A4: Classification (20%). Collocates shared across three registers are dominated by A13: Degree (14.29%), while collocates shared across two registers and NEWS-specific collocates both show their largest proportions in Q2: Speech (13.04 and 14.14%, respectively). Register-specific tendencies are also observed: CONV-specific collocates are most strongly associated with S4: Kin (20.83%), PARL-specific collocates with A7: Probability (9.8%), and ACAD-specific collocates with G2: Crime, law, and order (20%).

Taken together, the results indicate that partial sharing and register-specific patterns are more prominent than complete cross-register overlap. The following section examines these patterns in greater detail by focusing on the relational collocations of *say* according to their degree of cross-register sharing and register specificity.

Table 23

Distribution of Shared and Register-Specific Relational Collocations of Say across Registers

Collocate group	n	Example	Dominant category		
			Grammatical relation	Semantic domain	Semantic subdomain
Shared (4 registers)	5	do	X and/or ... (40%)	Q: Language and communication (40%)	A4 Classification (20%)
Shared (3 registers)	21	have	Modifiers of X (42.86%)	A: General and abstract terms (38.1%)	A13: Degree (14.29%)
Shared (2 registers)	69	minister	Subjects of X (49.28%)	Q: Language and communication (21.74%)	Q2: Speech (13.04%)
CONV-specific	48	reply	Subjects of X (75%)	S: Social actions, states, and processes (37.5%)	S4: Kin (20.83%)
PARL-specific	51	gently	Subjects of X (43.14%)	A: General and abstract terms (25.49%)	A7: Probability (9.8%)
NEWS-specific	97	source	X and/or (43.43%)	Q: Language and communication (17.35%)	Q2: Speech (14.14%)
ACAD-specific	30	constitute	Subjects of X (56.67%)	A: General and abstract terms (30%)	G2: Crime, law, and order (20%)

Note. n = number of collocate types. Dominant grammatical relations, semantic domains, and semantic subdomains are based on the highest number of distinct collocate types within each group.

Relational collocations shared across four and three registers

Table 24 and Table 25 present a closer view of the relational collocations of *say* that are shared across four and three registers, respectively. The two tables focus on the specific collocate type

involved, the register combinations in which they occur, their grammatical relations, and their semantic domains. Information about their association strength and more detailed semantic analysis is provided in **Appendix E2**.

Table 24 lists the five relational collocations of *say* attested across all four registers. These collocates form a small cross-register core, suggesting that only a limited set of relational collocations is shared by CONV, PARL, NEWS, and ACAD. They are not restricted to a single grammatical environment, but distributed across three grammatical relations. *Do* and *write* occur in the coordinated verbal patterns *X and/or*, *exactly* functions as modifiers of *say*, and *thing* and *word* occur as objects of *say*.

Semantically, the modifier *exactly* belongs to A: General and abstract terms and contributes to specifying or emphasizing the manner and accuracy of saying. The object collocates *thing* and *word* represent general content or linguistic units associated with what is said, while the coordinated collocates *do* and *write* relate to actions and forms of communication.

In addition, although all five collocates occur in the four registers, their association strength varies. An examination of the LogDice range column shows that some collocates, such as *word* (6.34–10.50), vary considerably in collocational strength, whereas others, such as *write* (8.21–9.22), remain relatively stable.

Table 24

Relational Collocations of Say Shared across Four Registers

Collocate	Grammatical relation	Registers attested	Log Dice range	Semantic domain
do	X and/or	CONV, PARL, NEWS, ACAD	8.16–11.25	Z: Names and grammar
write	X and/or	CONV, PARL, NEWS, ACAD	8.21–9.22	Q: Language and communication
exactly	Modifiers of X	CONV, PARL, NEWS, ACAD	5.29–7.32	A: General and abstract terms
thing	Objects of X	CONV, PARL, NEWS, ACAD	5.67–9.80	O: Substances, materials, objects, and equipment
word	Objects of X	CONV, PARL, NEWS, ACAD	6.34–10.50	Q: Language and communication

Note. Log Dice range refers to the minimum and maximum Log Dice values observed across the registers in which the collocate is attested.

Table 25 shows that relational collocations shared across three registers are more numerous and more varied in terms of lexical types and grammatical environments in comparison with the fully shared group. The collocates occur across four grammatical relations, with *objects of X* not represented. Regarding the register combinations, two patterns of cross-register overlap are identified. The larger group consists of collocates shared by CONV, PARL, and NEWS (13 types). This group includes a relatively wide range of items, from movement-related collocates such as *come* and *go*, to stance modifiers such as *probably* and *really*, and subject collocates such as *people*.

By contrast, the PARL-NEWS-ACAD group contains fewer collocates (8 types), including *think* (*X and/or*), *have* (*infinitive objects of X*), *explicitly*, *merely*, *rightly*, and *simply* (*modifiers of X*), as well as *judge* and *officer* (*subjects of X*). Compared with the CONV-PARL-NEWS group, this group is more restricted in size and contains items that are more closely associated with abstract, evaluative, or institutional contexts. Notably, no collocates are shared by the

CONV-PARL-ACAD or CONV-NEWS-ACAD combinations, indicating limited collocational overlap in register groupings that include both CONV and ACAD.

Overall, the fully shared collocates of *say* form a restricted core characterized by general meanings and basic communicative functions, while three-register collocates display greater lexical, grammatical, and semantic diversity. The comparison between the CONV-PARL-NEWS and PARL-NEWS-ACAD groups indicates that cross-register overlap is uneven: the former contains a wider range of collocates associated with spoken and public discourse, while the latter represents a more limited set of collocations linked to abstract, evaluative, and institutional contexts.

Table 25

Relational Collocations of Say Shared across Three Registers

Collocate	Grammatical relation	Semantic domain
<i>Register combination: CONV–PARL–NEWS</i>		
come	X and/or	M: Movement, location, travel, and transport
go	X and/or	M: Movement, location, travel, and transport
stop	X and/or	T: Time
basically	Modifiers of X	Z: Names and grammar
honestly	Modifiers of X	Z: Names and grammar
only	Modifiers of X	A: General and abstract terms
probably	Modifiers of X	A: General and abstract terms
really	Modifiers of X	A: General and abstract terms
government	Subjects of X	G: Government and public
letter	Subjects of X	Q: Language and communication
people	Subjects of X	S: Social actions, states, and processes
police	Subjects of X	G: Government and public
woman	Subjects of X	S: Social actions, states, and processes
<i>Register combination: PARL–NEWS–ACAD</i>		
think	X and/or	X: Psychological actions, states, and processes
have	Infinitive objects of X	A: General and abstract terms
explicitly	Modifiers of X	A: General and abstract terms
merely	Modifiers of X	A: General and abstract terms
rightly	Modifiers of X	A: General and abstract terms
simply	Modifiers of X	A: General and abstract terms
judge	Subjects of X	G: Government and public
officer	Subjects of X	G: Government and public

Relational Collocations shared across two registers

Table 26 presents the relational collocations of *say* shared across the six possible two-register combinations. A comparison across the six register pairs shows that two-register overlap is highly uneven. The strongest overlap occurs between PARL and NEWS, which share 39 collocates. This accounts for more than half of all collocates shared across two registers, indicating a strong collocational overlap between the two public or institutional registers. The remaining register pairs contain considerably fewer shared collocates: CONV-NEWS has 11, followed by

CONV-PARL (5). Overlaps involving ACAD are particularly limited, with only 6 collocates shared by CONV-ACAD, 5 by PARL-ACAD, and 3 by NEWS-ACAD.

Table 26 shows that collocates shared between CONV and the other three registers are mainly concentrated in *subjects of X*, including items such as *person* (CONV-ACAD), *mother* and *man* (CONV-NEWS), and *colleague* (CONV-PARL). The CONV-NEWS subject collocates are particularly associated with S: Social actions, states, and processes, highlighting references to social participants. This pair also shows greater grammatical variety, with shared collocates occurring across four grammatical relations, including *X and/or* (*ring*, *try*), *modifiers of X* (*suddenly*), and *objects of X* (*number*). This suggests that CONV shares a wider range of collocational patterns with NEWS than with ACAD or PARL.

The much larger PARL-NEWS pair is also more grammatically diverse. The shared collocates occur across three grammatical relations: *X and/or*, *modifiers of X*, and *subjects of X*. A notable feature of this pair is the presence of many public and institutional terms. Examples of subject collocates include *authority*, *company*, *council*, and *expert*. It also includes collocates related to communication and evaluation, such as *point*, *important*, *clearly*, and *publicly*. Semantically, these collocates cover several domains, particularly general and abstract meanings, language and communication, government and public affairs, and social processes. This pattern suggests that the PARL-NEWS overlap reflects shared concerns with public actors, institutional topics, and reported statements, as well as a relatively strong orientation towards spoken or speech-based content in both registers.

Overall, the extent of two-register overlap varies considerably across register pairs. PARL and NEWS share by far the largest number of relational collocations of *say*, whereas the other register pairs show much smaller sets of shared collocates. Pairs involving ACAD contain particularly few items. This pattern, however, should be interpreted in light of the much lower raw frequency of *say* in ACAD, which may partly limit the number of collocates available for cross-register sharing.

Table 26

Relational Collocations of Say Shared across Two Registers

Collocate	Grammatical relation	Semantic domain
<i>Register combination: CONV-ACAD</i>		
sentence	Objects of X	Q: Language and communication
head	Subjects of X	B: The body and the individual
person	Subjects of X	S: Social actions, states, and processes
policeman	Subjects of X	G: Government and public
teacher	Subjects of X	P: Education
text	Subjects of X	Q: Language and communication
<i>Register combination: CONV-NEWS</i>		
ring	X and/or	Q: Language and communication
try	X and/or	X: Psychological actions, states, and processes
suddenly	Modifiers of X	N: Numbers and measurement
number	Objects of X	N: Numbers and measurement
boss	Subjects of X	S: Social actions, states, and processes
doctor	Subjects of X	B: The body and the individual
friend	Subjects of X	S: Social actions, states, and processes

Continued on next page

Table 26 continued

Collocate	Grammatical relation	Semantic domain
group	Subjects of X	S: Social actions, states, and processes
man	Subjects of X	S: Social actions, states, and processes
mother	Subjects of X	S: Social actions, states, and processes
paper	Subjects of X	Q: Language and communication
<i>Register combination: CONV–PARL</i>		
name	Objects of X	Q: Language and communication
prayer	Objects of X	S: Social actions, states, and processes
colleague	Subjects of X	I: Money and commerce in industry
note	Subjects of X	Q: Language and communication
parent	Subjects of X	S: Social actions, states, and processes
<i>Register combination: NEWS–ACAD</i>		
involve	Infinitive objects of X	A: General and abstract terms
represent	Infinitive objects of X	E; Q: Emotion; language and communication
critic	Subjects of X	Q: Language and communication
<i>Register combination: PARL–ACAD</i>		
specifically	Modifiers of X	A: General and abstract terms
Court	Subjects of X	G: Government and public
Secretary	Subjects of X	I: Money and commerce in industry
respondent	Subjects of X	Q: Language and communication
speaker	Subjects of X	Q: Language and communication
<i>Register combination: PARL–NEWS</i>		
agree	X and/or	A: General and abstract terms
apologise	X and/or	Q: Language and communication
laugh	X and/or	E: Emotion
lie	X and/or	A: General and abstract terms
look	X and/or	A: General and abstract terms
mean	X and/or	Q: Language and communication
pause	X and/or	T: Time
point	X and/or	Q: Language and communication
put	X and/or	M: Movement, location, travel, and transport
see	X and/or	X: Psychological actions, states, and processes
shrug	X and/or	X: Psychological actions, states, and processes
smile	X and/or	E: Emotion
stand	X and/or	M: Movement, location, travel, and transport
almost	Modifiers of X	A: General and abstract terms
clearly	Modifiers of X	A: General and abstract terms
confidently	Modifiers of X	E: Emotion
consistently	Modifiers of X	A: General and abstract terms
famously	Modifiers of X	X: Psychological actions, states, and processes
immediately	Modifiers of X	N: Numbers and measurement
previously	Modifiers of X	N: Numbers and measurement
publicly	Modifiers of X	A: General and abstract terms

Continued on next page

Table 26 continued

Collocate	Grammatical relation	Semantic domain
repeatedly	Modifiers of X	N: Numbers and measurement
safely	Modifiers of X	A: General and abstract terms
well	Modifiers of X	A: General and abstract terms
Association	Subjects of X	S: Social actions, states, and processes
House	Subjects of X	H: Architecture, housing, and the home
Labour	Subjects of X	G: Government and public
Office	Subjects of X	I: Money and commerce in industry
authority	Subjects of X	G: Government and public
committee	Subjects of X	S: Social actions, states, and processes
company	Subjects of X	I: Money and commerce in industry
council	Subjects of X	G: Government and public
expert	Subjects of X	X: Psychological actions, states, and processes
leader	Subjects of X	S: Social actions, states, and processes
minister	Subjects of X	G: Government and public
report	Subjects of X	Q: Language and communication
spokesman	Subjects of X	Q: Language and communication
statement	Subjects of X	Q: Language and communication

Register-specific relational collocations

Tables 27–30 present the register-specific relational collocations of *say* across four grammatical relations: *subjects of X*, *objects of X*, *modifiers of X*, and *infinitive objects of X*. For each relation, the tables report the collocates specific to each register, together with their dominant semantic domains and subdomains. No register-specific collocates were found for *X and/or*. Collectively, these tables show that register-specific collocations are unevenly distributed across both registers and grammatical relations, reflecting distinct grammatical and semantic preferences in the use of *say*.

Among the four grammatical relations, *subjects of X* contains the largest number of register-specific collocates across the four registers, as shown in Table 27. CONV has the highest number of register-specific subject collocates (36), followed by NEWS (31), PARL (22), and ACAD (17). The dominant semantic domain differs by register: CONV is characterized by a concentration of collocates in S: Social actions, states, and processes, with S4 Kin representing the most prominent collocates, such as *brother*, *grandma*, *grandpa*, *granny*, *mammy*, *mum*, *sister*, and *wife*. In contrast, PARL is mainly associated with Q: Language and communication (e.g., *party*, *opposition*, and *speech*), whereas NEWS and ACAD show a strong orientation towards G: Government and public. This contrast suggests that register-specific subject collocates reflect different semantic patterns across registers: CONV emphasizes social participants, PARL foregrounds communicative roles, and NEWS and ACAD are more closely linked to institutional actors.

As shown in Table 28, register-specific object collocates are relatively few compared with subject collocates. CONV contains the largest number, with 6 collocates, while PARL and NEWS have only one each, and ACAD has two. The relatively small number of register-specific object collocates can be considered against the valency patterns of *say*. Herbst et al. (2004) describe *say* as “making a statement, either in speech or in writing” (p. 720). Importantly, the content associated with *say* can be realised through several complement types, including an NP, a that-clause, a wh-clause, and quoted or sentential material, while further complements may be

Table 27*Register-Specific Subject Collocates of Say across Registers*

Register	<i>n</i>	Collocates	Dominant domain	Dominant subdomain
CONV	36	bloke, book, boy, brother, card, class, computer, dad, email, girl, grandad, grandma, grandpa, granny, guy, kid, lad, lady, lecturer, mammy, mate, message, mum, nan, nurse, player, question, side, sign, sister, solicitor, staff, student, thing, voice, wife	S: Social actions, states, and processes; 18/36 (50.00%)	S4: Kin; 10/36 (27.78%)
PARL	22	amendment, assessment, bench, Bill, business, Chancellor, Commission, constituent, country, document, Foundation, guidance, manifesto, member, Memorandum, opposition, Paper, Parliament, party, speech, Treasury, Trust	Q: Language and communication; 5/22 (22.73%)	G1: Government and politics; 4/22 (18.18%)
NEWS	31	agency, analyst, bank, campaigner, charity, chief, commission, department, diplomat, executive, family, firm, insider, lawyer, manager, ministry, neighbour, office, official, president, prosecutor, regulator, researcher, scientist, secretary, source, spokesperson, spokeswoman, star, union, witness	G: Government and public; 7/31 (22.58%)	G1: Government and politics; 4/31 (12.90%)
ACAD	17	Appeal, author, child, clause, counsel, court, defendant, Department, father, helper, law, philosopher, plaintiff, reader, statute, therapist, writer	G: Government and public; 6/17 (35.29%)	G2: Crime, law, and order; 6/17 (35.29%)

Note. *n* = number of collocates.

introduced by *to* and *about*. Nominal objects therefore constitute only one possible realisation of the complementation of *say*. Against this broader valency pattern, the register-specific object collocates can be examined in terms of their lexical and semantic distribution across registers.

The CONV-specific object collocates include socially situated event *birthday*, abstract and explanatory content nouns *bit*, *fact*, *idea*, *reason*, as well as the informal vague-content noun *stuff*. The pattern suggests that, in conversation, *say* takes object collocates that are closely related to everyday content, explanation, and vague or non-specific reference. The remaining registers display more limited but semantically distinctive collocational patterns. PARL has only *opposite*, associated with contrast and opposing positions in the context of parliamentary debates, while NEWS has *problem*, an abstract issue-related noun. In ACAD, the two object collocates, *syllable* and *utterance*, belong to Q: Language and communication, reflecting a more specialized focus on linguistic units.

Table 28*Register-Specific Object Collocates of Say across Registers*

Register	<i>n</i>	Collocates	Dominant domain	Dominant subdomain
CONV	6	birthday, bit, fact, idea, reason, stuff	A: General and abstract terms; 2/6 (33.33%)	A2: Affect; 1/6 (16.67%)
PARL	1	opposite	M: Movement, location, travel, and transport; 1/1 (100.00%)	M6: Location and direction; 1/1 (100.00%)
NEWS	1	problem	A: General and abstract terms; 1/1 (100.00%)	A12: Easy/difficult; 1/1 (100.00%)

Continued on next page

Table 28*Register-Specific Object Collocates of Say across Registers*

Register	<i>n</i>	Collocates	Dominant domain	Dominant subdomain
ACAD	2	syllable, utterance	Q: Language and communication; 2/2 (100.00%)	Q1: Linguistic actions, states, and processes; Communication; 1/2 (50.00%)

Note. *n* = number of collocates.

Turning to modification patterns, Table 29 shows that register-specific *modifiers of X* are most frequent in PARL (17), followed by NEWS (14), CONV (4), and ACAD (4). Across all four registers, A: General and abstract terms is the dominant semantic domain, but the subdomains differ. CONV and PARL are both associated with A7: Probability, NEWS with E4: Happiness and contentment, and ACAD with A14: Exclusivizers/particularizers. This suggests that register-specific modifiers are broadly abstract, but differ in the kinds of meaning they encode.

Table 29*Register-Specific Modifier Collocates of Say across Registers*

Register	<i>n</i>	Collocates	Dominant domain	Dominant subdomain
CONV	4	apparently, definitely, maybe, obviously	A: General and abstract terms; 3/4 (75.00%)	A7: Probability; 2/4 (50.00%)
PARL	17	briefly, categorically, certainly, constantly, correctly, effectively, eloquently, essentially, frequently, gently, kindly, openly, perhaps, politely, precisely, respectfully, seriously	A: General and abstract terms; 8/17 (47.06%)	A7 Probability: 3/17 (17.65%)
NEWS	14	allegedly, bluntly, cheerfully, firmly, initially, loud, nearly, pointedly, privately, proudly, quietly, reportedly, ruefully, sadly	A: General and abstract terms; 5/14 (35.71%)	E4: Happiness and contentment; 4/14 (28.57%)
ACAD	4	above, expressly, truly, typically	A: General and abstract terms; 3/4 (75.00%)	A14: Exclusivizers/particularizers; 1/4 (25.00%)

Note. *n* = number of collocates.

Finally, Table 30 presents register-specific infinitive object collocates, which are limited in number but occur across all four registers. The distribution shows a stronger representation in NEWS (10) and ACAD (7), compared with PARL (5) and especially CONV (1), suggesting that infinitive-object patterns are more characteristic of written registers. Semantically, these collocates are dominated by A: General and abstract terms, although the specific meanings differ across registers. In PARL, the collocates include verbs *assure*, *convince*, *explain*, *reassure*, and *reflect*, which are linked to explanation and persuasion. NEWS includes a broader set of abstract verbs, such as *believe*, *include*, *show*, and *want*, reflecting reported attitudes, inclusion, and presentation of information. In ACAD, collocates such as *constitute*, *exist*, *indicate*, *lead*, *lie*, *occur*, and *provide* are more strongly associated with abstract description, existence, causation, and argumentation. By contrast, the single CONV-specific collocate, *do*, reflects the use of *say* with a general action verb, which is consistent with the tendency of conversation to refer to everyday activities and action.

Table 30*Register-Specific Infinitive Object Collocates of Say across Registers*

Register	<i>n</i>	Collocates	Dominant domain	Dominant subdomain
CONV	1	do	A: General and abstract terms; 1/1 (100.00%)	A1: General and abstract terms; 1/1 (100.00%)
PARL	5	assure, convince, explain, reassure, reflect	A: General and abstract terms; 3/5 (60.00%)	A7: Probability; 2/5 (40.00%)
NEWS	10	believe, contain, enjoy, favour, help, include, own, resemble, show, want	A: General and abstract terms; 4/10 (40.00%)	A1: General and abstract terms; 2/10 (20.00%)
ACAD	7	constitute, exist, indicate, lead, lie, occur, provide	A: General and abstract terms; 6/7 (85.71%)	A3: Being; 2/7 (28.57%)

Note. *n* = number of collocates.

Taken together, Tables 27–30 demonstrate that register-specific collocates of *say* are shaped by clear interactions between grammatical relations and semantic preferences. *Subjects of X* is the most widely distributed relation and contains the most register-specific collocates, whereas the other grammatical relations display more distinct register preferences: *objects of X* are mainly represented in CONV, *modifiers of X* are most frequent in PARL, and *infinitive objects of X* are more prominent in written registers. Semantically, CONV is characterized by interpersonal and everyday meanings, reflected in subject collocates referring to social participants, especially kinship terms, and object collocates related to personal events, general information, and vague reference. PARL shows patterns associated with communicative, political, and institutional contexts through subject collocates, together with evaluative meanings in modifier collocates. NEWS highlights institutional meanings through subject collocates referring to public actors. ACAD displays a stronger orientation towards abstract and specialized meanings, particularly evident in infinitive-object collocates related to description, existence, causation, and argumentation, as well as object collocates associated with linguistic concepts.

4.3.5 Interim summary

The case study of *say* examined how *say* varies across registers in terms of grammatical behaviour, validated relational collocations, patterns of cross-register sharing, and semantic preference. An initial examination of automatically extracted grammatical relations indicated that register differences are relation-specific rather than uniform. CONV is strongly associated with pronominal patterns, PARL shows a more mixed profile involving prepositional-phrase and particle constructions, NEWS is characterized by object- and particle-related patterns, and ACAD is distinguished by infinitival and coordinated patterns. Subsequent manual validation of the Word Sketch output highlighted the need to verify automatically extracted data before collocational analysis. Many candidate entries did not correspond to the grammatical relations assigned by the parser, with passive-use data being particularly unreliable. Therefore, subsequent analyses were based solely on validated relational collocations, improving the reliability of the dataset.

Building on this validated dataset, the analysis of relational collocations further pointed to register-based differences in both the overall distribution of collocates and the ways in which they are distributed across the five grammatical relations. NEWS contained the largest number of

distinct collocates, suggesting the greatest lexical variety, whereas PARL showed the highest normalized proportion of relational collocations relative to the total occurrences of *say*. Across grammatical relations, *subjects of X* and *modifiers of X* were the most frequent relations, while *X and/or* was less common. In addition to these overall distributional tendencies, each register showed preferences for particular grammatical relations: PARL was strongly associated with subject, modifier, and object collocates; NEWS was marked by prominent subject collocates, but showed more distinctive infinitive-object patterns; ACAD was most clearly characterized by infinitive objects and coordinated constructions; and CONV was mainly distinguished by the use of object collocates.

A comparison of shared and register-specific collocates showed that the collocational profile of *say* is shaped more by register-dependent lexical selection and varying degrees of collocational overlap than by a stable set of collocates shared across all registers. Fully shared collocates were limited, while two-register, three-register, and register-specific collocates were more numerous. The strongest overlap occurred between PARL and NEWS, consistent with their shared public, institutional, and speech-oriented characteristics. By contrast, fewer collocates overlapped between ACAD and the other registers, a pattern that may partly reflect the lower frequency of *say* in ACAD.

Finally, the semantic analysis provided further evidence that register variation in the relational collocations of *say* emerges through the interaction of grammatical relation, lexical choice, and semantic preference. The examination of semantic domains and subdomains showed that shared collocates tend to represent relatively general meanings connected with the basic communicative functions of *say*, whereas register-specific collocates reveal more context-dependent semantic patterns. CONV was associated with interpersonal and everyday meanings, PARL with communicative and evaluative meanings in public argumentation, NEWS with institutional actors and evaluative reporting, and ACAD with abstract, specialized, and argument-oriented meanings.

In sum, these findings show that relational collocations provide a useful perspective for examining how grammatical structure, lexical selection, and meaning interact across registers. The analysis of six grammatical relations revealed clear differences across registers, although no valid relational collocations were identified for *complements of X*. Since the case study is based on a single high-frequency lexical verb, the patterns observed for *say* cannot be assumed to represent broader verb behaviour. The following section therefore extends the analysis to the 25 high-frequency lexical verbs to examine whether comparable grammatical, distributional, and semantic patterns emerge on a broader scale.

4.4 Distribution of Relational Collocations across Registers and Three Selected Grammatical Relations for the 25 Verbs

4.4.1 Introduction

For the cross-verb analysis, three grammatical relations are examined: *objects of X*, *complements of X*, and *modifiers of X*. *Objects of X* were selected because verb + direct object combinations have been extensively examined in learner-corpus research and have been shown to pose difficulties for L2 learners and to distinguish advanced proficiency levels, as discussed in **Section 2.2.4.2**. *Complements of X* were included to incorporate verb complementation, while *modifiers of X* were selected to broaden the scope to patterns of verbal modification. To address the first and second research questions, **Section 4.4.2** examines the overall distribution of relational collocations across registers, providing a broad comparison of their frequency and occurrence across the 25 verbs. **Section 4.4.3** then compares their distribution across the three grammatical relations,

revealing how the observed register differences vary by grammatical relation.

4.4.2 Overall distribution of relational collocations across registers and the three grammatical relations

Before examining the distributional results, all candidate relational collocations for the 25 lexical verbs across the four registers and three selected grammatical relations were manually validated. In total, 14,503 candidate entries were checked, of which 9,088 were retained, 5,405 were excluded, and 10 were merged as variants. The results reported below are based on this validated dataset, which is provided in **Appendix E1**.

Table 31 summarizes the overall distribution of relational collocations across the 25 lexical verbs by register. NEWS records the largest number of distinct collocates (3,502) and the highest total raw frequency of relational collocations (3,020,884), followed by PARL, CONV, and ACAD. However, the 25 verbs are also substantially more frequent in NEWS than in the other registers, increasing the opportunities for grammatical relations and collocations to occur, as reported in Table 14. The raw counts should therefore be interpreted in relation to overall verb frequency. When the frequency of relational collocations is calculated as a proportion of all occurrences of the 25 verbs in each register, PARL shows the highest proportion of relational collocations (23.28%), closely followed by ACAD (22.28%), while NEWS is lower (15.12%) and CONV has the lowest proportion (6.88%). Thus, although NEWS contains the largest absolute number of relational collocations and distinct collocates, these patterns partly reflect the much higher frequency of the verbs in this register; proportionally, relational collocations are more prominent in PARL and ACAD.

Table 31

Overall Distribution of Relational Collocations of the 25 Lexical Verbs by Register

Register	Total no. of collocates	Total raw frequency		Percentage
		Relational collocations	25 lexical verbs	
CONV	1,353	46,731	679,668	6.88%
PARL	2,674	740,807	3,182,354	23.28%
NEWS	3,502	3,020,884	19,973,814	15.12%
ACAD	1,559	60,785	272,845	22.28%

Note. Percentages were calculated using the total number of occurrences of the 25 lexical verbs in each register as the denominator.

Table 32 presents the verb-level distribution of validated relational collocations across the four registers. The total number of relational collocations varies considerably across verbs, ranging from 123 for *try* to 789 for *make*. These cross-verb differences should, however, be interpreted in relation to valency patterns of individual verbs, since lexical verbs differ in the grammatical patterns in which they typically occur. For the present register comparison, the percentages are informative because they indicate how frequently each verb occurs with relational collocations relative to its overall frequency in each register. Substantial cross-register variation is evident for many verbs. For example, the proportion for *make* ranges from 23.94% in CONV to 51.57% in PARL, with intermediate values in NEWS (37.51%) and ACAD (42.88%).

Across the 25 lexical verbs, the register with the highest proportion of relational collocations varies by verb. PARL records the highest value for 16 of the 25 verbs, whereas NEWS is highest for *go*, *know*, *come*, *look*, and *try*, ACAD for *give*, *think*, and *show*, and CONV for *want*. These

highest percentages are highlighted in bold in Table 32. This distribution shows that the relative prominence of relational collocations is not associated with the same register across all verbs, but varies according to the individual verb.

Table 32

Verb-Level Distribution of Relational Collocations across Registers

Verb	CONV	PARL	NEWS	ACAD	Total
say	23 (2.00%)	45 (4.27%)	41 (0.97%)	15 (1.61%)	124
make	125 (23.94%)	208 (51.57%)	237 (37.51%)	219 (42.88%)	789
take	98 (19.90%)	134 (52.13%)	179 (38.61%)	114 (46.91%)	525
see	80 (8.24%)	124 (20.50%)	179 (15.47%)	86 (19.96%)	469
go	66 (2.78%)	94 (5.76%)	164 (8.68%)	23 (4.67%)	347
know	57 (4.15%)	82 (8.29%)	107 (10.20%)	42 (9.15%)	288
give	44 (5.54%)	103 (34.19%)	122 (20.79%)	113 (36.46%)	382
come	12 (0.95%)	34 (1.44%)	59 (3.90%)	19 (0.03%)	124
think	11 (1.41%)	63 (3.70%)	133 (3.35%)	25 (4.57%)	232
use	69 (6.11%)	155 (33.74%)	147 (21.52%)	142 (30.22%)	513
get	178 (13.49%)	198 (31.29%)	227 (22.58%)	53 (14.53%)	656
need	69 (8.88%)	116 (17.24%)	114 (14.76%)	80 (14.08%)	379
work	27 (6.88%)	64 (21.75%)	80 (14.95%)	25 (7.39%)	196
find	71 (13.85%)	212 (35.12%)	235 (23.54%)	133 (21.84%)	651
look	53 (7.84%)	81 (7.32%)	186 (11.44%)	15 (4.66%)	335
leave	48 (9.86%)	137 (28.05%)	192 (20.23%)	74 (19.91%)	451
put	96 (14.70%)	128 (28.13%)	140 (20.70%)	45 (15.03%)	409
mean	23 (2.46%)	104 (9.79%)	67 (7.72%)	28 (5.93%)	222
want	88 (12.93%)	101 (3.79%)	137 (3.47%)	11 (0.49%)	337
pay	28 (11.11%)	91 (54.47%)	121 (31.65%)	56 (36.91%)	296
call	33 (5.05%)	93 (14.36%)	176 (9.46%)	33 (6.50%)	335
ask	13 (6.33%)	65 (29.10%)	115 (14.03%)	35 (22.57%)	228
show	6 (1.97%)	122 (20.71%)	147 (17.16%)	118 (20.92%)	393
try	14 (2.99%)	34 (2.29%)	69 (4.05%)	6 (1.48%)	123
bring	19 (6.25%)	83 (18.50%)	119 (15.18%)	47 (15.74%)	268

Note. Values in the register columns are presented as n (%), where n refers to the number of validated relational collocates for the verb in the corresponding register. Percentages are calculated by dividing the total raw frequency of these validated relational collocates by the total number of occurrences of the verb in that register. The Total column reports the total number of validated relational collocates across the four registers.

4.4.3 Distribution of relational collocations by grammatical relation and register

Tables 33–35 provide a more detailed view of how the relational collocations of the 25 verbs are distributed across registers within *objects of X*, *modifiers of X*, and *complements of X*. The verb collocations occurring in these three relations show different cross-register distributions. Across the three selected grammatical relations, for most verbs, object collocations generally account for the larger proportion of verb occurrences than modifier or complement collocations, although the pattern varies by verb and register.

For object collocates, most verbs have their highest proportion in PARL, as shown in Table 33. The exceptions show different register preferences: *say*, *go* and *want* have their highest proportions in CONV; *know*, *work*, and *try* in NEWS; and *give* and *show* in ACAD.

Modifier collocates show greater variation across verbs, as shown in Table 34: twelve verbs have their highest proportions in PARL, six in NEWS (*go*, *come*, *think*, *get*, *leave*, *pay*), four in CONV (*make*, *need*, *want*, *try*), and three in ACAD (*see*, *know*, *use*). Thus, unlike object

collocations, modifier collocations do not show similarly strong concentration in one register.

Complement collocates show a more restricted and uneven distribution across the verb set, as illustrated in Table 35. They are absent or occur at very low proportions for several verbs, including *say*, *know*, *give*, *use*, *need*, *mean*, *pay*, and *try*. Among verbs with more substantial complement collocations, the register in which the highest proportion occurs also varies, although NEWS contains the largest group of such verbs, including *take*, *go*, *come*, *get*, *look*, *put*, *want*, and *call*. In comparison, PARL is more prominent for *make*, *see*, *find*, *pay*, and *show*, while ACAD shows the highest proportion for *think*, *leave*, and *bring*.

These differences should be considered in relation to the valency patterns of individual verbs, which constrain the grammatical constructions available to them. However, valency alone does not account for the observed cross-register variation, since the same verb may show different distributions across registers depending on the grammatical relation. This is particularly clear for *leave*: object collocations reach their highest proportion in PARL (25.53%), modifier collocations in NEWS (1.38%), and complement collocations in ACAD (2.64%). A similar relation-specific pattern is found for *make*, whose object (41.24%) and complement (8.73%) collocations are proportionally highest in PARL, while its modifier collocations are highest in CONV (1.70%). The concordance lines in Example (4) illustrate these within-verb differences, providing further evidence of register-related variation in verb collocations.

(4)	Register- and relation-specific uses of <i>leave</i> and <i>make</i>
a.	We have seen over the years people leaving other countries in the European Union as a result of anti-Semitic incidents. (<i>leave + object, PARL</i>)
b.	He has until February to leave voluntarily and has already made inquiries about somewhere to live in Britain. (<i>leave + modifier, NEWS</i>)
c.	We must allow, too, that this argument as it stands leaves our understanding untouched . (<i>leave + complement, ACAD</i>)
d.	It is sometimes described as having to demonstrate that a decision made as part of a ministerial discretion is arbitrary, perverse or capricious. (<i>make + object, PARL</i>)
e.	I have previously made Ministers aware that there are no beds for females in Dorset who need intensive psychiatric care. (<i>make + complement, PARL</i>)
f.	do it dad it'll make you happier mm (<i>make + modifier, CONV</i>)

Table 36 summarises, for each verb, the grammatical relations in which relational collocations are significantly overused in a target register relative to one or more of the other three registers. Full statistical results are provided in **Appendix D**. The extent of significant overuse varies considerably across registers. PARL and NEWS each account for 44 verb–relation cases of significant overuse, compared with 24 in ACAD and only eight in CONV. Thus, significant cross-register differences are more widely distributed across the verb set in PARL and NEWS than in ACAD and CONV.

The strongest contrasts, in which the same grammatical relation is significantly overused in one register relative to all three others, are restricted to a smaller set of verbs. In CONV, this pattern is found for the object collocations of *go*. In PARL, modifier collocations are register-distinctively overused for *look*, *put*, and *say*. NEWS shows two such patterns: complement collocations for *call*, *come*, and *take*, and object collocations for *try* and *work*. In ACAD, the complement collocations of *think* are overused relative to all three other registers. The results provide evidence that some high-frequency lexical verbs show register-specific lexico-grammatical preferences,

with the same verb displaying different patterns of overuse in object, complement, and modifier relations across registers.

Among these register-distinctive cases, *go* and *try* are particularly noteworthy because both show significant overuse in all four registers, but the grammatical relation involved differs by register. To illustrate this pattern, Example (5) presents concordance lines for *try*, with each concordance line corresponding to a grammatical relation that is significantly overused in the target register relative to the comparison register(s) shown in Table 36.

(5)	Concordance examples of register-distinctive overuse of <i>try</i>
a.	it's useful to go and try stuff on cos you know when to get it in the sale (<i>CONV — modifiers of X, overused relative to ACAD</i>)
b.	We know that the ombudsman tried desperately to get timely responses from the Treasury and other Departments, and was frustrated at every turn. (<i>PARL — modifiers of X, overused relative to ACAD</i>)
c.	A powerful forward with 14 goals this season, he is ready to try his luck in a bigger league. (<i>NEWS — objects of X, overused relative to CONV, PARL, and ACAD</i>)
d.	David Lammy, the Labour MP for Tottenham, has been trying hard to publicise this issue and is highly vocal in calling for black fathers to be more involved with their children. (<i>NEWS — modifiers of X, overused relative to ACAD</i>)
e.	The above methods are however well tried , possess adequate theoretical bases and in general, are simple tests, that is, they test only one quantity and are not derived tests which combined several variables. (<i>ACAD — objects of X, overused relative to PARL</i>)

Broader relation-specific tendencies are also evident in the pairwise comparisons. Relative to CONV, object collocations are significantly overused for 12 verbs in PARL, 13 in NEWS, and 10 in ACAD. Modifier collocations show the strongest tendency in PARL, where significant overuse is identified for 14 verbs, while complement collocations are most prominent in NEWS, involving 15 verbs. These inferential results broadly support the descriptive patterns in Tables 33–35 and further show that the observed register differences vary across verbs and grammatical relations.

Overall, the results reveal that register-related variation in the relational collocations of the 25 verbs is both verb-specific and relation-specific. The same grammatical relation may show different patterns of overuse across verbs and registers, while individual verbs may display different register preferences across object, modifier, and complement relations. Having established these distributional and inferential differences, the next section examines the extent to which the relational collocations themselves are shared across registers or are specific to particular registers.

Table 33*Distribution of Object Collocates across Registers for the 25 Verbs*

Verb	CONV	PARL	NEWS	ACAD	Total
say	12 (1.08%)	5 (1.07%)	4 (0.39%)	5 (0.81%)	26
make	78 (18.43%)	87 (41.24%)	95 (28.77%)	100 (34.48%)	360
take	83 (17.56%)	86 (47.73%)	98 (35.89%)	88 (44.21%)	355
see	67 (6.66%)	83 (18.73%)	88 (13.55%)	61 (17.41%)	299
go	4 (0.05%)	—	—	—	4
know	49 (2.22%)	66 (4.04%)	81 (5.43%)	26 (3.48%)	222
give	39 (4.68%)	72 (33.02%)	86 (19.70%)	91 (35.35%)	288
use	59 (4.63%)	100 (29.88%)	104 (18.98%)	94 (25.25%)	357
get	75 (8.75%)	92 (24.76%)	88 (13.85%)	41 (9.95%)	296
need	59 (4.73%)	94 (13.11%)	95 (10.78%)	72 (11.44%)	320
work	6 (0.93%)	10 (0.67%)	21 (2.16%)	3 (0.42%)	40
find	49 (8.19%)	89 (24.04%)	94 (16.55%)	95 (15.62%)	327
look	—	2 (0.04%)	1 (0.03%)	—	3
leave	41 (8.60%)	63 (25.53%)	73 (16.23%)	61 (16.79%)	238
put	82 (13.54%)	94 (23.95%)	91 (18.67%)	41 (13.95%)	308
mean	3 (0.73%)	80 (6.52%)	40 (5.15%)	18 (2.68%)	141
want	79 (7.29%)	83 (2.74%)	86 (2.36%)	8 (0.23%)	256
pay	25 (10.48%)	55 (51.64%)	69 (28.67%)	50 (35.57%)	199
call	33 (5.05%)	59 (12.09%)	54 (8.02%)	27 (5.51%)	173
ask	12 (6.23%)	38 (26.90%)	84 (12.78%)	33 (22.27%)	167
show	6 (1.97%)	95 (17.98%)	89 (15.57%)	100 (18.65%)	290
try	3 (0.41%)	11 (0.39%)	40 (1.68%)	3 (0.79%)	57
bring	19 (6.25%)	65 (17.49%)	89 (14.08%)	42 (14.92%)	215

Note. Values are presented as n (%), where n refers to the number of validated relational collocates attested for the given verb, register, and grammatical relation. Percentages are calculated by dividing the total raw frequency of the validated collocates in that grammatical relation by the total number of occurrences of the verb in the corresponding register. The Total column reports the total number of validated collocates in the relevant grammatical relation across the four registers. A dash indicates that no validated collocate was attested.

Table 34*Distribution of Modifier Collocates across Registers for the 25 Verbs*

Verb	CONV	PARL	NEWS	ACAD	Total
say	11 (0.92%)	40 (3.20%)	37 (0.58%)	10 (0.81%)	98
make	14 (1.70%)	42 (1.60%)	45 (1.22%)	29 (1.35%)	130
take	15 (2.34%)	47 (4.39%)	67 (2.58%)	26 (2.70%)	155
see	13 (1.58%)	28 (1.37%)	31 (1.54%)	24 (2.45%)	96
go	29 (2.10%)	36 (3.58%)	56 (4.46%)	12 (2.70%)	133
know	8 (1.93%)	15 (4.25%)	26 (4.76%)	16 (5.68%)	65
give	5 (0.85%)	31 (1.17%)	35 (1.09%)	22 (1.11%)	93
come	10 (0.91%)	23 (1.20%)	32 (2.66%)	17 (0.03%)	82
think	11 (1.41%)	23 (2.53%)	32 (2.68%)	13 (1.75%)	79
use	10 (1.48%)	55 (3.86%)	42 (2.54%)	48 (4.97%)	155
get	18 (1.71%)	23 (1.09%)	26 (1.76%)	—	67
need	10 (4.16%)	22 (4.13%)	15 (3.97%)	8 (2.64%)	55
work	21 (5.95%)	51 (21.00%)	54 (12.71%)	22 (6.98%)	148
find	4 (1.02%)	23 (1.46%)	31 (1.46%)	18 (2.33%)	76
look	14 (2.67%)	35 (6.44%)	31 (2.65%)	9 (3.14%)	89
leave	5 (0.84%)	31 (1.04%)	35 (1.38%)	4 (0.47%)	75
put	14 (1.16%)	28 (3.35%)	25 (1.19%)	4 (1.08%)	71
mean	20 (1.74%)	24 (3.27%)	26 (2.56%)	10 (3.26%)	80
want	8 (5.62%)	17 (1.05%)	14 (1.07%)	3 (0.26%)	42
pay	3 (0.62%)	35 (2.81%)	52 (2.97%)	6 (1.33%)	96
call	—	28 (2.13%)	35 (1.10%)	6 (0.99%)	69
ask	1 (0.10%)	27 (2.20%)	31 (1.25%)	2 (0.30%)	61
show	—	22 (2.64%)	24 (1.47%)	17 (2.23%)	63
try	11 (2.58%)	23 (1.90%)	28 (2.37%)	3 (0.69%)	65
bring	—	15 (0.87%)	21 (0.76%)	3 (0.46%)	39

Table 35*Distribution of Complement Collocates across Registers for the 25 Verbs*

Verb	CONV	PARL	NEWS	ACAD	Total
say	—	—	—	—	0
make	33 (3.81%)	79 (8.73%)	97 (7.52%)	91 (7.04%)	300
take	—	1 (0.01%)	15 (0.13%)	—	16
see	—	13 (0.40%)	60 (0.37%)	1 (0.10%)	74
go	33 (0.63%)	58 (2.18%)	108 (4.22%)	11 (1.97%)	210
know	—	1 (0.00%)	—	—	1
give	—	—	1 (0.00%)	—	1
come	2 (0.04%)	11 (0.24%)	27 (1.24%)	2 (0.00%)	42
think	—	40 (1.16%)	102 (0.67%)	12 (2.82%)	154
use	—	—	1 (0.00%)	—	1
get	87 (3.03%)	85 (5.44%)	116 (6.97%)	13 (4.58%)	301
need	—	—	4 (0.01%)	—	4
work	—	3 (0.08%)	5 (0.08%)	—	8
find	18 (4.65%)	100 (9.62%)	110 (5.53%)	20 (3.88%)	248
look	39 (5.17%)	44 (0.84%)	155 (8.75%)	6 (1.53%)	244
leave	2 (0.42%)	43 (1.48%)	85 (2.62%)	9 (2.64%)	139
put	—	7 (0.83%)	24 (0.84%)	—	31
mean	—	—	1 (0.00%)	—	1
want	1 (0.03%)	1 (0.00%)	37 (0.04%)	—	39
pay	—	1 (0.02%)	1 (0.00%)	—	2
call	—	6 (0.14%)	87 (0.34%)	—	93
show	—	5 (0.09%)	34 (0.12%)	1 (0.05%)	40
try	—	—	1 (0.00%)	—	1
bring	—	3 (0.14%)	10 (0.34%)	2 (0.36%)	15

Table 36*Register-Distinctive Overuse of Grammatical Relations across Verbs*

Target register and grammatical relation	Overused relative to	Verbs	<i>n</i>
<i>CONV</i>			
objects of X	PARL; NEWS; ACAD	<i>go</i>	1
complements of X	PARL; ACAD	look	1
modifiers of X	ACAD	get, <i>try</i>	2
objects of X	ACAD	want, work	2
modifiers of X	PARL	come	1
objects of X	NEWS	say	1
<i>PARL</i>			
modifiers of X	CONV; NEWS; ACAD	look, put, say	3
complements of X	CONV; ACAD	call, find, put, see, work	5
modifiers of X	CONV; ACAD	ask, call, pay, work	4
objects of X	CONV; ACAD	call, get, mean	3
objects of X	CONV; NEWS	ask	1
objects of X	CONV	bring, find, give, leave, look, make, need, pay, see, show, take, use	12
complements of X	CONV	bring, <i>go</i> , leave, make, think	5
modifiers of X	CONV	bring, know, show, use	4
modifiers of X	ACAD	get, leave, <i>try</i>	3
complements of X	NEWS	know, pay	2
objects of X	ACAD	want	1
objects of X	NEWS	say	1
<i>NEWS</i>			
complements of X	CONV; PARL; ACAD	call, come, take	3
objects of X	CONV; PARL; ACAD	<i>try</i> , work	2
complements of X	CONV; ACAD	<i>go</i> , put, see, work	4
complements of X	CONV; PARL	bring, want	2
modifiers of X	CONV; ACAD	ask, pay	2
complements of X	PARL; ACAD	look	1
modifiers of X	PARL; ACAD	come	1
objects of X	CONV	ask, bring, find, give, know, look, mean, need, pay, see, show, take, use	13
modifiers of X	CONV	bring, call, <i>go</i> , know, show, work	6
complements of X	CONV	get, leave, think	3
modifiers of X	ACAD	get, leave, <i>try</i>	3
complements of X	PARL	give, need	2
complements of X	ACAD	show	1
objects of X	ACAD	want	1
<i>ACAD</i>			
complements of X	CONV; PARL; NEWS	think	1
complements of X	CONV; PARL	bring	1
objects of X	CONV	ask, bring, give, mean, need, pay, see, show, take, use	10
modifiers of X	CONV	bring, call, find, know, pay, show, use	7
complements of X	CONV	<i>go</i> , leave, see	3
objects of X	NEWS	say	1
objects of X	PARL	<i>try</i>	1

Note. Overused grammatical relations were identified using a Bonferroni-adjusted significance threshold of $p < .0083$ and $|\text{Log Ratio}| > 1$. n = number of verbs.

4.5 Shared and Register-Specific Relational Collocations across the 25 Verbs

4.5.1 Introduction

Section 4.5 first presents the overall distribution of shared and register-specific relational collocations, followed by analyses of collocations shared across all four, three, and two registers, and those specific to individual registers. The final subsection examines the semantic patterns of relational collocations in the object relation across the 25 verbs. The full lists of collocates and their semantic classifications are provided in **Appendix E3** and **E4**, respectively.

4.5.2 Overall distribution of shared and register-specific collocations

Table 37 presents the overall distribution of shared and register-specific relational collocations across the 25 verbs. The results show that relational collocations shared by all four registers are limited, with only 218 items. By contrast, collocations shared by two registers are much more frequent (1,265) than those shared by three registers (697). Register-specific collocations account for a substantial proportion of the total, with NEWS showing the largest number of register-specific items (1,508), followed by PARL (905), ACAD (630), and CONV (552).

At the level of individual verbs, the extent of sharing and specificity varies considerably. *Make* shows the greatest degree of cross-register sharing, with the largest number of collocations shared across all four registers (31) and across two-register combinations (111). *Get*, by contrast, records the highest number of collocations shared across three registers (75). At the other end of the scale, *think* and *bring* have no relational collocations shared across all four registers. Register-specific collocations are also unevenly distributed across verbs, with the highest number found for *put* in CONV (64), *find* in PARL (72), *call* in NEWS (112), and *make* in ACAD (82).

To examine these patterns in greater detail, the following sections move from the most widely shared collocations to increasingly register-specific ones, beginning with relational collocations shared across four registers, followed by those shared across three and two registers, and finally those specific to individual registers.

Table 37

Overall Distribution of Shared and Register-Specific Relational Collocations across the 25 Verbs

Verb	Shared 4	Shared 3	Shared 2	CONV- specific	PARL- specific	NEWS- specific	ACAD- specific	Total
say	3	9	18	10	18	15	6	79
make	31	71	111	44	46	59	82	444
take	26	51	49	40	29	64	38	297
see	9	34	63	35	44	82	44	311
go	9	30	61	15	19	64	1	199
know	7	23	45	18	27	42	14	176
give	16	31	43	15	30	47	47	229
come	1	10	22	5	6	28	7	79
think	—	20	38	1	14	77	5	155
use	14	50	60	30	51	45	61	311
get	23	75	71	61	66	66	12	374
need	5	36	39	44	48	41	40	253
work	7	12	32	10	23	30	5	119
find	15	51	96	19	72	85	70	408

Continued on next page

Table 37*Overall Distribution of Shared and Register-Specific Relational Collocations across the 25 Verbs*

Verb	Shared 4	Shared 3	Shared 2	CONV- specific	PARL- specific	NEWS- specific	ACAD- specific	Total
look	2	21	56	12	27	111	3	232
leave	8	27	70	21	51	101	26	304
put	9	25	55	64	52	60	13	278
mean	5	16	33	7	54	19	8	142
want	3	11	49	59	52	77	6	257
pay	16	25	39	5	19	45	11	160
call	2	16	48	13	41	112	17	249
ask	3	5	50	5	25	60	11	159
show	3	22	58	1	47	71	80	282
try	1	6	23	4	9	40	2	85
bring	—	20	36	14	35	67	21	193
Total	218	697	1,265	552	905	1,508	630	5,775

4.5.3 Relational collocations shared across all four registers

Table 38 presents the relational collocations shared across all four registers. In general, these items are unevenly distributed across verbs and grammatical relations. Object collocates form the largest group, particularly with high-frequency and semantically general verbs such as *take*, *make*, *give*, *pay*, *get*, and *use*. Among these verbs, *take* has the largest number of fully shared object collocates, with 23 items, including *action*, *advantage*, *care*, *child*, *control*, *drug*, *hour*, *interest*, *look*, *minute*, *month*, *note*, *notice*, *part*, *place*, *responsibility*, *risk*, *route*, *seat*, *step*, *time*, *view*, and *year*. Similarly, *make*, *give*, and *pay* each have 16 object collocates, while *get* has 11 and *use* has 10.

Complement collocates also constitute an important part of the fully shared set, but they are concentrated in a smaller number of verbs. This pattern is observed for *make*, which has 14 complement collocates, including *angry*, *aware*, *bad*, *cheap*, *clear*, *different*, *difficult*, *easy*, *good*, *hard*, *possible*, *redundant*, *simple*, and *sure*. It is also evident for *get*, which has 12 complement collocates, including *bad*, *big*, *drunk*, *free*, *high*, *married*, *old*, *rich*, *rid*, *right*, *small*, and *wrong*. By contrast, modifier collocates are generally fewer and more dispersed, although they occur with verbs such as *go* (e.g., *abroad*, *deep*, *fast*, *further*, *outside*, *straight*) and *work* (e.g., *closely*, *full-time*, *hard*, *properly*, *together*, *well*).

Overall, the most stable relational collocations across the four registers are mainly realised through object relations, with complement relations forming a more verb-specific pattern.

Table 38*Relational Collocates Shared across All Four Registers by Verb and Grammatical Relation*

Verb	Grammatical relation	Collocates	<i>n</i>
say	Objects of X	thing; word	2
	Modifiers of X	exactly	1
	Complements of X	—	0
make	Objects of X	appointment; arrangement; change; choice; comment; decision; difference; effort; life; mistake; point; profit; progress; sense; statement; use	16

Continued on next page

Table 38*Relational Collocates Shared across All Four Registers by Verb and Grammatical Relation*

Verb	Grammatical relation	Collocates	<i>n</i>
take	Modifiers of X	probably	1
	Complements of X	angry; aware; bad; cheap; clear; different; difficult; easy; good; hard; possible; redundant; simple; sure	14
	Objects of X	action; advantage; care; child; control; drug; hour; interest; look; minute; month; note; notice; part; place; responsibility; risk; route; seat; step; time; view; year	23
	Modifiers of X	long; normally; seriously	3
	Complements of X	—	0
see	Objects of X	child; difference; light; problem; reason; thing; woman; work	8
	Modifiers of X	only	1
	Complements of X	—	0
go	Objects of X	—	0
	Modifiers of X	abroad; deep; fast; further; outside; straight	6
	Complements of X	bankrupt; mad; wrong	3
know	Objects of X	answer; detail; fact; name; rule	5
	Modifiers of X	exactly; well	2
	Complements of X	—	0
give	Objects of X	advice; answer; birth; choice; detail; example; impression; information; instruction; notice; number; opportunity; permission; reason; time; way	16
	Modifiers of X	—	0
	Complements of X	—	0
come	Objects of X	—	0
	Modifiers of X	late	1
	Complements of X	—	0
think	Objects of X	—	0
	Modifiers of X	—	0
	Complements of X	—	0
use	Objects of X	datum; form; language; phrase; power; skill; technology; term; tool; word	10
	Modifiers of X	commonly; normally; only; regularly	4
	Complements of X	—	0
get	Objects of X	chance; child; deal; hand; hold; idea; impression; job; message; money; right	11
	Modifiers of X	—	0
	Complements of X	bad; big; drunk; free; high; married; old; rich; rid; right; small; wrong	12
need	Objects of X	help	1
	Modifiers of X	desperately; necessarily; obviously; only	4
	Complements of X	—	0
work	Objects of X	shift	1
	Modifiers of X	closely; full-time; hard; properly; together; well	6
	Complements of X	—	0
find	Objects of X	answer; job; place; solution; way; word; work	7
	Modifiers of X	only	1

Continued on next page

Table 38*Relational Collocates Shared across All Four Registers by Verb and Grammatical Relation*

Verb	Grammatical relation	Collocates	<i>n</i>
look	Complements of X	attractive; difficult; easy; hard; helpful; interesting; useful	7
	Objects of X	—	0
	Modifiers of X	—	0
	Complements of X	different; similar	2
leave	Objects of X	child; country; home; room; school	5
	Modifiers of X	early	1
	Complements of X	alone; open	2
put	Objects of X	child; effort; finger; hand; name; point; pressure; weight; word	9
	Modifiers of X	—	0
	Complements of X	—	0
mean	Objects of X	thing	1
	Modifiers of X	literally; necessarily; presumably; really	4
	Complements of X	—	0
want	Objects of X	child	1
	Modifiers of X	desperately; really	2
	Complements of X	—	0
pay	Objects of X	amount; attention; bill; debt; fee; fine; interest; money; premium; price; rate; rent; salary; sum; tax; wage	16
	Modifiers of X	—	0
	Complements of X	—	0
call	Objects of X	election; police	2
	Modifiers of X	—	0
	Complements of X	—	0
ask	Objects of X	parent; people; question	3
	Modifiers of X	—	0
	Complements of X	—	0
show	Objects of X	interest; picture; sign	3
	Modifiers of X	—	0
	Complements of X	—	0
try	Objects of X	—	0
	Modifiers of X	hard	1
	Complements of X	—	0
bring	Objects of X	—	0
	Modifiers of X	—	0
	Complements of X	—	0

4.5.4 Relational collocations shared across three registers

Table 39 presents relational collocations shared across three registers, grouped by register combination and grammatical relation. On the whole, the register combinations vary in both the number of collocates attested and the word combinations they attract. For *objects of X* and *modifiers of X*, all four possible register combinations are attested, whereas *complements of X* occur in three combinations, with CONV–PARL–ACAD absent.

For *objects of X*, the strongest overlap is found in PARL–NEWS–ACAD, with 225 collocates, followed by CONV–PARL–NEWS with 116. By contrast, combinations that include

both CONV and ACAD contain far fewer object collocates, with 25 in CONV-NEWS-ACAD and only one in CONV-PARL-ACAD. A similar pattern can be observed for *modifiers of X*. The largest number occurs in PARL-NEWS-ACAD (106), followed by CONV-PARL-NEWS (87), while CONV-NEWS-ACAD and CONV-PARL-ACAD each contain only one modifier collocate. A contrasting distribution emerges for *complements of X*. The highest count appears in CONV-PARL-NEWS (74), followed by PARL-NEWS-ACAD (58), whereas CONV-NEWS-ACAD includes only three complement collocates.

These patterns suggest that collocations shared across registers are most common when PARL and NEWS are part of the combination, regardless of the grammatical relation. In contrast, combinations that include CONV and ACAD tend to have far fewer shared collocates across objects, modifiers, and complements. Such distributional tendencies can be further illustrated by the representative collocates in the final column in Table 39, which show how three-register sharing is realised linguistically across the three grammatical relations. Object collocates include combinations such as *make + contact* in CONV-NEWS-ACAD, *make + announcement* in CONV-PARL-NEWS, and *make + allegation* in PARL-NEWS-ACAD. Modifier collocates include combinations such as *think + genuinely* in CONV-PARL-NEWS and *think + carefully* in PARL-NEWS-ACAD. Complement collocates are represented by combinations such as *go + blind* in CONV-PARL-NEWS and *go + public* in PARL-NEWS-ACAD. These examples further illustrate variation in the collocates shared by different register combinations within the same grammatical relation. Although the same verb may occur across registers, the specific collocations with which it co-occurs differ across register combinations, reflecting differences in communicative purposes.

Table 39

Relational Collocates Shared across Three Registers by Register Combination and Grammatical Relation

Register combination	<i>n</i>	Verbs involved	Representative collocates
<i>Panel A. Objects of X</i>			
CONV-NEWS-ACAD	25	make; take; see; give; use; get; find; leave; put; want; pay; bring	make + contact; take + photograph; see + doctor; give + address; use + computer; get + feeling; find + food; leave + space; put + mark; want + help; pay + deposit; bring + child
CONV-PARL-ACAD	1	find	find + example
CONV-PARL-NEWS	116	make; take; see; know; give; use; get; need; work; find; leave; put; want; pay; call; ask; show; bring	make + announcement; take + credit; see + end; know + area; give + chance; use + Internet; get + food; need + answer; work + way; find + body; leave + door; put + arm; want + change; pay + bonus; call + ambulance; ask + permission; show + face; bring + people
PARL-NEWS-ACAD	225	make; take; see; know; give; use; get; need; work; find; leave; put; mean; pay; call; ask; show; try; bring	make + allegation; take + account; see + case; know + extent; give + access; use + device; get + access; need + access; work + land; find + accommodation; leave + area; put + case; mean + end; pay + benefit; call + halt; ask + authority; show + ability; try + case; bring + action
<i>Panel B. Modifiers of X</i>			
CONV-NEWS-ACAD	1	take	take + literally
CONV-PARL-ACAD	1	make	make + only

Continued on next page

Table 39

Relational Collocates Shared across Three Registers by Register Combination and Grammatical Relation

Register combination	<i>n</i>	Verbs involved	Representative collocates
CONV–PARL–NEWS	87	say; make; take; see; go; know; give; come; think; get; need; work; find; look; leave; put; mean; want; try	say + basically; make + almost; take + certainly; see + exactly; go + backwards; know + probably; give + only; come + probably; think + genuinely; get + almost; need + certainly; work + abroad; find + probably; look + online; leave + late; put + online; mean + basically; want + necessarily; try + constantly
PARL–NEWS–ACAD	106	say; make; take; see; go; know; give; come; think; use; need; work; find; look; leave; put; mean; pay; call; ask; show; try; bring	say + explicitly; make + necessarily; take + further; see + above; go + directly; know + commonly; give + freely; come + close; think + carefully; use + currently; need + badly; work + actively; find + commonly; look + backwards; leave + deliberately; put + simply; mean + automatically; pay + badly; call + commonly; ask + specifically; show + clearly; try + deliberately; bring + inevitably
<i>Panel C. Complements of X</i>			
CONV–NEWS–ACAD	3	make	make + comfortable
CONV–PARL–NEWS	74	make; go; come; get; find; look	make + happy; go + blind; come + free; get + angry; find + awkward; look + attractive
PARL–NEWS–ACAD	58	make; see; go; come; think; get; find; look; leave; show; bring	make + acceptable; see + fit; go + public; come + alive; think + appropriate; get + large; find + able; look + promising; leave + free; show + willing; bring + close

4.5.5 Relational collocations shared across two registers

Table 40 presents relational collocations shared across two registers, categorised by register pair and grammatical relation. The distribution reveals that these shared relational collocations are unevenly distributed across the register combinations, with the most extensive overlap consistently occurring in pairs involving NEWS and PARL.

With respect to *objects of X*, all six register pairs are attested, with PARL–NEWS being the predominant category, yielding 291 shared collocates. This substantial overlap is followed by CONV–NEWS (129), PARL–ACAD (106), and NEWS–ACAD (81). In comparison, object collocates shared by CONV–PARL (18) and CONV–ACAD (13) are considerably more limited.

A similar distributional pattern is also observed for *modifiers of X*. PARL–NEWS again has the highest number of shared modifier collocates with 230 collocates, far exceeding the other register pairs. CONV–NEWS and NEWS–ACAD each contain 41 modifier collocates, followed by PARL–ACAD with 20. The register combinations CONV–PARL (11) and CONV–ACAD (3) similarly represent the least frequent pairings for shared collocates.

By contrast, the distribution of *complements of X* is more restricted, with only five register pairs attested and the complete absence of the CONV–ACAD pairing. In a similar vein to the other two grammatical relations, PARL–NEWS retains the largest share of complement collocates (194), followed by CONV–NEWS (59). The pairings involving ACAD exhibit far lower frequencies, with NEWS–ACAD and PARL–ACAD yielding only 15 and 12 collocates, respectively. Finally, the overlap within the CONV–PARL pair is marginal, containing only a single attested collocate (*i.e.*, *come + true*).

Examples in the final column help clarify the lexical characteristics associated with these two-register overlaps. For instance, in object relations, PARL–NEWS includes combinations such as *make + commitment*, *give + assurance*, and *ask + Commission*, which point to institutional, political, and public-discourse contexts. In modifier relations, the same register pair contains manner- and stance-related combinations, such as *put + bluntly* and *use + correctly*. Complement relations show a similar evaluative tendency, with combinations such as *see + able*, *think + acceptable*, and *find + abhorrent*, where the complement contributes to judgement and assessment. These examples suggest that the extensive PARL–NEWS overlap is not only quantitative, but also reflects the lexical choices associated with the shared communicative concerns of the two registers, particularly institutional and public affairs, together with the expression of stance and evaluation.

Table 40

Relational Collocates Shared across Two Registers by Register Pair and Grammatical Relation

Register pair	<i>n</i>	Verbs involved	Representative collocates
<i>Panel A. Objects of X</i>			
CONV–ACAD	13	say; make; take; see; know; give; use; leave; put; ask	say + sentence; make + connection; take + tablet; see + pattern; know + language; give + idea; use + dictionary; leave + baby; put + piece; ask + person
CONV–NEWS	129	say; make; take; see; know; give; use; get; need; work; find; leave; put; want; pay; call; ask; show; try; bring	say + number; make + deal; take + break; see + bit; know + boy; give + job; use + hand; get + call; need + break; work + job; find + bit; leave + car; put + face; want + baby; pay + dollar; call + boy; ask + advice; show + emotion; try + thing; bring + food
CONV–PARL	18	say; make; take; know; use; get; find; leave; put; want; pay; call; bring	say + name; make + time; take + word; know + amount; use + English; get + amount; find + mechanism; leave + thing; put + date; want + person; pay + loan; call + place; bring + thing
NEWS–ACAD	81	make; take; see; know; give; use; need; work; find; leave; put; pay; call; ask; show; try; bring	make + appearance; take + sample; see + page; know + cause; give + account; use + code; need + encouragement; work + miracle; find + expression; leave + city; put + blame; pay + commission; call + conference; ask + audience; show + change; try + method; bring + change
PARL–ACAD	106	make; take; see; know; give; use; get; need; find; leave; put; mean; want; pay; call; ask; show; bring	make + adjustment; take + evidence; see + advantage; know + boundary; give + assistance; use + approach; get + compensation; need + assessment; find + argument; leave + care; put + argument; mean + care; want + student; pay + grant; call + attention; ask + Secretary; show + degree; bring + law
PARL–NEWS	291	make; take; see; know; give; use; get; need; work; find; look; leave; put; mean; want; pay; call; ask; show; try; bring	make + commitment; take + comfort; see + action; know + bound; give + assurance; use + experience; get + act; need + action; work + hand; find + alternative; look + horse; leave + House; put + amount; mean + amount; want + access; pay + farmer; call + bluff; ask + Commission; show + appreciation; try + approach; bring + attention
<i>Panel B. Modifiers of X</i>			
CONV–ACAD	3	make; see; work	make + normally; see + especially; work + backwards

Continued on next page

Table 40*Relational Collocates Shared across Two Registers by Register Pair and Grammatical Relation*

Register pair	<i>n</i>	Verbs involved	Representative collocates
CONV–NEWS	41	say; make; take; see; go; know; think; use; get; work; look; put; mean; want; try	say + suddenly; make + definitely; take + definitely; see + barely; go + completely; know + only; think + definitely; use + probably; get + definitely; work + fine; look + definitely; put + outside; mean + absolutely; want + definitely; try + actually
CONV–PARL	11	see; know; get; look; leave; put; want; try	see + well; know + obviously; get + absolutely; look + obviously; leave + literally; put + only; want + only; try + basically
NEWS–ACAD	41	make; take; see; go; know; give; come; think; use; need; work; find; look; put; pay; call; show	make + largely; take + initially; see + below; go + astray; know + formerly; give + initially; come + ashore; think + generally; use + mostly; need + scarcely; work + alone; find + occasionally; look + rather; put + hard; pay + fully; call + aptly; show + below
PARL–ACAD	20	say; make; take; see; come; use; find; look; leave; ask; show	say + specifically; make + directly; take + indeed; see + precisely; come + inevitably; use + appropriately; find + consistently; look + briefly; leave + thereby; ask + legitimately; show + graphically
PARL–NEWS	230	say; make; take; see; go; know; give; come; think; use; get; need; work; find; look; leave; put; mean; want; pay; call; ask; show; try; bring	say + almost; make + absolutely; take + almost; see + certainly; go + alone; know + certainly; give + certainly; come + certainly; think + certainly; use + correctly; get + barely; need + simply; work + collaboratively; find + certainly; look + actively; leave + effectively; put + bluntly; mean + ultimately; want + certainly; pay + currently; call + affectionately; ask + constantly; show + apparently; try + actively; bring + certainly
<i>Panel C. Complements of X</i>			
CONV–NEWS	59	make; go; get; find; look; want	make + big; go + bad; get + boring; find + boring; look + amazing; want + big
CONV–PARL	1	come	come + true
NEWS–ACAD	15	make; go; come; think; look; leave	make + profitable; go + berserk; come + near; think + unnecessary; look + alike; leave + intact
PARL–ACAD	12	make; think; find	make + appropriate; think + fit; find + satisfactory
PARL–NEWS	194	make; see; go; come; think; get; work; find; look; leave; put; want; call; show; bring	make + affordable; see + able; go + AWOL; come + cheap; think + acceptable; get + active; work + flat; find + abhorrent; look + bleak; leave + able; put + good; want + open; call + civilised; show + able; bring + alive

4.5.6 Register-specific relational collocations

The distribution of register-specific relational collocations is detailed in Table 41. While these collocations are most prevalent in object relations, their distribution varies considerably across both registers and grammatical relations.

With respect to *objects of X*, NEWS records the largest number of register-specific collocates (563), followed closely by PARL (532), ACAD (468), and CONV (441). Modifier collocates are less frequent than object collocates, but they show a similar distributional tendency. NEWS again yields the highest count (307), followed by PARL (251), ACAD (122), and CONV (73). For both object and modifier relations, NEWS and PARL show higher counts than ACAD and CONV. The

pattern is different for *complements of X*, where the distribution is much more uneven across registers. NEWS shows an overwhelming predominance with 638 register-specific collocates, substantially exceeding PARL (122), ACAD (40), and CONV (38). This sharp asymmetry indicates that NEWS is characterized by the distinctive use of complement relations.

As shown in the final column, the representative collocates illustrate how register specificity extends beyond distributional differences to include lexical and semantic variation across the three grammatical relations. In object relations, ACAD includes items related to specialized or discipline-specific contexts, such as *use + algorithm*, *find + abnormality*, and *ask + academics*, whereas CONV contains more everyday and interactional word combinations, such as *say + birthday*, *make + bed*, and *call + aunt*. NEWS shows a broader range of combinations referring to general public issues or widely relevant topics, such as *say + problem*, *need + aid*, *want + advice*, and *leave + army*, while PARL contains explicitly parliamentary or legislative combinations such as *say + opposite*, *make/take + amendment*, and *ask + Chancellor*.

This lexical contrast is equally apparent when examining the modifier collocates. ACAD is associated with formal or textual modifiers such as *think + commonly*, *find + experimentally*, *call + hereinafter*, and *try + summarily*, while CONV favours informal stance markers such as *make/take/pay/ask/work/use + basically*. NEWS includes evidential and reportative modifiers such as *say/make/take/use/find/pay/show/try + allegedly*, and *put/call + accidentally*, whereas PARL contains procedural and temporal modifiers such as *say + briefly*, *find/get + currently*, and *bring + immediately*. Meanwhile, complement collocates are especially distinctive in NEWS, where combinations such as *make + anxious*, *think + absurd*, and *mean + unlikely* point to attitudinal and result-state descriptions.

In sum, the findings in **Sections 4.5** demonstrate that register variation is reflected in both the distribution of shared and register-specific relational collocations and the lexical items occurring within different grammatical relations. The object relation is selected for more detailed semantic analysis because it constitutes the largest category across both shared and register-specific categories and, as shown in the preceding distributional and inferential analysis, also displays clear register differences, particularly between CONV and the other three registers. The following section therefore examines the semantic patterns of object relational collocations across the 25 verbs.

Table 41

Register-Specific Collocates by Register and Grammatical Relation

Register	<i>n</i>	Verbs involved	Representative collocates
<i>Panel A. Objects of X</i>			
ACAD only	468	say; make; take; see; know; give; use; get; need; find; leave; put; mean; want; pay; call; ask; show; try; bring	say + syllable; make + admission; take + argument; see + Act; know + God; give + appearance; use + algorithm; get + coverage; need + adjustment; find + abnormality; leave + basis; put + asset; mean + age; want + measure; pay + allowance; call + Parliament; ask + academics; show + abnormality; try + action; bring + appeal
CONV only	441	say; make; take; see; go; know; give; use; get; need; work; find; leave; put; mean; want; pay; call; ask; show; try; bring	say + birthday; make + bed; take + age; see + advert; go + cinema; know + bloke; give + card; use + ability; get + bag; need + bag; work + arse; find + article; leave + bag; put + address; mean + place; want + TV; pay + cheque; call + aunt; ask + dad; show + age; try + stuff; bring + book

Continued on next page

Table 41*Register-Specific Collocates by Register and Grammatical Relation*

Register	<i>n</i>	Verbs involved	Representative collocates
NEWS only	563	say; make; take; see; know; give; use; get; need; work; find; leave; put; mean; want; pay; call; ask; show; try; bring	say + problem; make + amends; take + aim; see + back (n.); know + actor; give + approval; use + French; get + back (n.); need + aid; work + arrangement; find + audience; leave + army; put + Labour; mean + bill; want + advice; pay + adviser; call + House; ask + God; show + card; try + Botox; bring + ball
PARL only	532	say; make; take; see; know; give; use; get; need; work; find; look; leave; put; mean; want; pay; call; ask; show; try; bring	say + opposite; make + amendment; take + amendment; see + Bill; know + background; give + certainty; use + Act; get + agreement; need + agreement; work + department; find + agreement; look + constituent; leave + Chair; put + agreement; mean + ability; want + Bill; pay + credit; call + Bill; ask + Chancellor; show + ambition; try + avenue; bring + Ministers
<i>Panel B. Modifiers of X</i>			
ACAD only	122	say; make; take; see; know; give; come; think; use; work; find; look; put; want; call; show; try; bring	say + above; make + commonly; take + approximately; see + commonly; know + accurately; give + above; come + frequently; think + commonly; use + alone; work + consistently; find + experimentally; look + remarkably; put + briefly; want + merely; call + hereinafter; show + above; try + summarily; bring + effectively
CONV only	73	say; make; take; see; go; know; give; come; think; use; get; need; work; find; look; put; mean; pay; ask; try	say + apparently; make + basically; take + basically; see + maybe; go + bang; know + maybe; give + basically; come + definitely; think + suddenly; use + basically; get + basically; need + almost; work + basically; find + definitely; look + actually; put + almost; mean + especially; pay + basically; ask + basically; try + definitely
NEWS only	307	say; make; take; see; go; know; give; come; think; use; get; work; find; look; leave; put; mean; want; pay; call; ask; show; try; bring	say + allegedly; make + allegedly; take + allegedly; see + almost; go + all-out; know + affectionately; give + allegedly; come + all ye faithful; think + highly; use + allegedly; get + forward; work + beautifully; find + allegedly; look + abroad; leave + above; put + accidentally; mean + apparently; want + badly; pay + allegedly; call + accidentally; ask + anxiously; show + allegedly; try + allegedly; bring + almost
PARL only	251	say; make; take; see; go; know; give; come; think; use; get; need; work; find; look; leave; put; mean; want; pay; call; ask; show; try; bring	say + briefly; make + ably; take + absolutely; see + absolutely; go + currently; know + colloquially; give + absolutely; come + largely; think + creatively; use + actively; get + currently; need + absolutely; work + assiduously; find + currently; look + afresh; leave + completely; put + ably; mean + genuinely; want + absolutely; pay + adequately; call + constantly; ask + certainly; show + alone; try + certainly; bring + immediately
<i>Panel C. Complements of X</i>			
ACAD only	40	make; go; think; find; look; leave	make + amenable; go + absent; think + sufficient; find + natural; look + blue; leave + unaffected

Continued on next page

Table 41*Register-Specific Collocates by Register and Grammatical Relation*

Register	<i>n</i>	Verbs involved	Representative collocates
CONV only	38	make; go; get; find; look	make + awkward; go + fine; get + Chinese; find + online; look + full
NEWS only	638	make; take; see; go; give; come; think; use; get; need; work; find; look; leave; put; mean; want; pay; call; show; try; bring	make + anxious; take + alive; see + akin; go + Dutch; give + large; come + big; think + absurd; use + raw; get + airborne; need + alive; work + free; find + abandoned; look + acceptable; leave + aghast; put + bad; mean + unlikely; want + able; pay + free; call + British; show + adept; try + free; bring + considerable
PARL only	122	make; take; see; go; know; come; think; get; work; find; look; leave; put; pay; call	make + challenging; take + amiss; see + possible; go + independent; know + available; come + new; think + helpful; get + accurate; work + sick; find + alarming; look + aghast; leave + bad; put + necessary; pay + direct; call + Muslim

4.5.7 Semantic analysis of relational collocations in the object relation across registers for the 25 verbs

Semantic profiles of shared object collocates

Table 42 presents the semantic domains of object collocates shared across four, three, and two registers. The results reveal that shared object collocates cover a wide range of semantic domains, but tend to concentrate in several key domains.

With regard to collocates shared across all four registers, three semantic domains are particularly prominent: A: General and abstract terms (16.43%), Q: Language and communication (16.43%), and X: Psychological actions, states and processes (15.71%). Together, these three domains account for nearly half of the four-register shared object collocates. This suggests that these collocates tend to involve general abstract meanings, communicative acts, and mental or cognitive processes. Representative examples include *get + chance/hold* in domain A, *ask + question* and *find + answer/word* in domain Q, and *get + idea/impression/message* in domain X.

In terms of collocates shared across three registers, the leading domains are S: Social actions, states and processes (18.53%), A: General and abstract terms (14.17%), I: Money and commerce in industry (11.72%), and X: Psychological actions, states and processes (10.90%). Collectively, these four domains make up slightly more than half of three-register shared object collocates. Compared with the four-register shared group, the three-register group gives greater prominence to social and economic meanings. Examples include *ask + member*, *bring + child/order/people* in domain S, *bring + action/case/proceeding* and *find + evidence/example* in domain A, *bring + charge*, *call + company* and *find + employment/money* in domain I, and *bring + experience/matter/perspective* in domain X.

A comparable, albeit more widely dispersed, distributional profile characterises the collocates shared across two registers. The five largest domains are A: General and abstract terms (15.36%), S: Social actions, states and processes (14.58%), X: Psychological actions, states and processes (11.76%), Q: Language and communication (9.56%), and I: Money and commerce in industry (8.62%). Together, these domains comprise approximately 59.88% of the two-register group. Examples include *bring + challenge/change/clarity/improvement/side/stability* in domain A, *ask + child/girl/man/mother/mum/panel/participant/person* in domain S, *bring + attention/expertise/hope/issue/news/skill* in domain X, and *ask + advice/reader/respondent* in

domain Q, and *ask + company/employer* in domain I.

Table 42

Semantic Domains of Shared Object Collocates

Semantic domain	No. of collocates (%)	Representative collocates
<i>Shared across 4 registers</i>		
A: General and abstract terms	23 (16.43%)	get + chance; get + hold; give + detail; give + example; give + opportunity; give + reason; know + detail; know + fact
Q: Language and communication	23 (16.43%)	ask + question; find + answer; find + word; give + advice; give + answer; give + instruction; give + notice; know + answer
X: Psychological actions, states and processes	22 (15.71%)	find + solution; find + way; get + idea; get + impression; get + message; give + choice; give + impression; give + information
S: Social actions, states and processes	16 (11.43%)	ask + parent; ask + people; get + child; get + right; give + permission; leave + child; need + help; put + child
I: Money and commerce in industry	15 (10.71%)	find + job; find + work; get + job; get + money; make + profit; pay + debt; pay + fee; pay + money
N: Numbers and measurement	8 (5.71%)	get + deal; give + number; pay + amount; pay + rate; pay + sum; put + pressure; put + weight; take + part
T: Time	7 (5.00%)	give + time; make + appointment; take + hour; take + minute; take + month; take + time; take + year
B: The body and the individual	5 (3.57%)	get + hand; give + birth; put + finger; put + hand; take + drug
O: Substances, materials, objects and equipment	4 (2.86%)	mean + thing; say + thing; see + thing; use + tool
G: Government, law and order	4 (2.86%)	call + election; call + police; know + rule; leave + country
M: Movement, location, travel and transport	4 (2.86%)	find + place; take + place; take + route; take + step
H: Architecture, housing and the home	3 (2.14%)	leave + home; leave + room; take + seat
P: Education	1 (0.71%)	leave + school
Z: Names and grammar	1 (0.71%)	pay + bill
L: Life and living things	1 (0.71%)	make + life
W: World and environment	1 (0.71%)	see + light
C: Arts and crafts	1 (0.71%)	show + picture
Y: Science and technology	1 (0.71%)	use + technology
<i>Shared across 3 registers</i>		
S: Social actions, states and processes	68 (18.53%)	ask + member; ask + permission; bring + benefit; bring + child; bring + order; bring + people; call + friend; call + man
A: General and abstract terms	52 (14.17%)	bring + action; bring + case; bring + proceeding; find + case; find + cause; find + evidence; find + example; find + fault
I: Money and commerce in industry	43 (11.72%)	ask + staff; bring + charge; call + company; call + strike; find + employment; find + money; get + credit; get + work
X: Psychological actions, states and processes	40 (10.90%)	bring + experience; bring + matter; bring + perspective; call + programme; call + system; find + information; get + feeling; get + information
Q: Language and communication	33 (8.99%)	bring + claim; call + book; get + answer; get + letter; get + response; give + name; give + sentence; give + speech
N: Numbers and measurement	31 (8.45%)	bring + pressure; find + level; find + space; get + number; give + figure; give + rise; give + weight; know + extent
O: Substances, materials, objects and equipment	17 (4.63%)	call + thing; find + thing; get + thing; know + thing; leave + mark; make + thing; pay + deposit; put + mark

Continued on next page

Table 42
Semantic Domains of Shared Object Collocates

Semantic domain	No. of collocates (%)	Representative collocates
M: Movement, location, travel and transport	17 (4.63%)	bring + end; find + path; find + site; get + access; give + access; know + area; leave + area; mean + end
B: The body and the individual	15 (4.09%)	call + ambulance; find + body; find + cure; get + head; get + treatment; leave + hospital; need + treatment; put + arm
T: Time	12 (3.27%)	call + halt; find + source; find + time; get + time; give + date; know + history; leave + time; make + start
H: Architecture, housing and the home	11 (3.00%)	find + accommodation; find + home; get + house; give + address; leave + door; leave + house; make + home; make + living
G: Government, law and order	7 (1.91%)	ask + authority; bring + prosecution; call + witness; get + vote; leave + government; use + law; use + weapon
E: Emotional actions, states and processes	5 (1.36%)	bring + peace; bring + relief; need + reassurance; take + pride; use + force
P: Education	4 (1.09%)	leave + university; need + training; take + test; use + test
F: Food and farming	3 (0.82%)	find + food; get + food; put + food
C: Arts and crafts	3 (0.82%)	get + picture; see + picture; take + photograph
K: Entertainment, sports and games	2 (0.54%)	call + tune; leave + party
W: World and environment	2 (0.54%)	see + world; work + land
Y: Science and technology	2 (0.54%)	use + Internet; use + computer
<i>Shared across 2 registers</i>		
A: General and abstract terms	98 (15.36%)	ask + Commission; ask + user; bring + challenge; bring + change; bring + clarity; bring + improvement; bring + side; bring + stability
S: Social actions, states and processes	93 (14.58%)	ask + child; ask + girl; ask + man; ask + mother; ask + mum; ask + panel; ask + participant; ask + person
X: Psychological actions, states and processes	75 (11.76%)	bring + attention; bring + expertise; bring + hope; bring + issue; bring + news; bring + skill; call + attention; call + expert
Q: Language and communication	61 (9.56%)	ask + advice; ask + reader; ask + respondent; call + film; call + helpline; call + name; call + report; find + argument
I: Money and commerce in industry	55 (8.62%)	ask + Secretary; ask + colleague; ask + company; ask + employer; bring + prosperity; find + buyer; find + role; find + staff
N: Numbers and measurement	40 (6.27%)	bring + number; bring + rate; bring + total; call + number; find + bit; find + deal; find + means; find + number
G: Government, law and order	38 (5.96%)	ask + council; ask + court; ask + government; ask + minister; ask + officer; ask + official; ask + police; bring + country
O: Substances, materials, objects and equipment	31 (4.86%)	bring + thing; bring + water; find + balance; find + comfort; find + mechanism; find + niche; find + spot; know + circumstance
M: Movement, location, travel and transport	30 (4.70%)	ask + driver; call + place; find + ground; find + route; get + place; give + direction; know + boundary; know + place
B: The body and the individual	26 (4.08%)	ask + patient; call + doctor; call + midwife; find + foot; get + vaccine; give + treatment; know + patient; leave + patient
E: Emotional actions, states and processes	22 (3.45%)	bring + joy; find + courage; find + peace; give + confidence; leave + force; leave + rest; make + joke; need + rest
T: Time	14 (2.19%)	call + time; get + day; get + minute; get + year; leave + baby; make + history; make + time; mean + delay
K: Entertainment, sports and games	13 (2.04%)	ask + audience; bring + party; call + song; leave + scene; make + music; need + break; put + party; see + game
P: Education	9 (1.41%)	ask + pupil; ask + student; ask + teacher; get + education; get + test; mean + test; need + education; want + school

Continued on next page

Table 42
Semantic Domains of Shared Object Collocates

Semantic domain	No. of collocates (%)	Representative collocates
L: Life and living things	9 (1.41%)	call + dog; find + life; give + life; know + life; look + horse; put + life; take + life; want + dog
H: Architecture, housing and the home	6 (0.94%)	ask + department; call + home; find + house; find + room; leave + House; want + home
F: Food and farming	5 (0.78%)	bring + food; leave + field; need + food; pay + farmer; put + kettle
W: World and environment	5 (0.78%)	bring + rain; know + nature; know + world; make + world; use + land
C: Arts and crafts	3 (0.47%)	take + photo; take + picture; want + picture
Y: Science and technology	3 (0.47%)	try + experiment; use + formula; use + software
Z: Names and grammar	2 (0.31%)	find + match; use + English

Note. Values are presented as *n* (%), with percentages calculated within each sharing type.

Semantic profiles of register-specific object collocates

Tables 43–46 display the semantic profiles of object collocates that are specific to individual registers. On the whole, the four registers show different semantic orientations, suggesting that register-specific object collocates are not only unevenly distributed in quantitative terms, but also differ in the kinds of meanings they encode.

For CONV-specific object collocates, Table 43 shows that the four largest semantic domains account for more than half of the total. These are F: Food and farming (18.14%), O: Substances, materials, objects and equipment (17.46%), B: The body and the individual (11.11%), and Q: Language and communication (10.66%). This profile points to a strong association with food-related items, everyday material objects, bodily references, and interpersonal communication. Examples include *bring + bread/wine* in domain F, *find + stuff/switch* in domain O, *get + eye/hair/leg* in domain B, and *get + e-mail/phone* in domain Q.

PARL-specific object collocates, illustrated in Table 44, display a different semantic profile. The four leading semantic domains are A: General and abstract terms (16.17%), S: Social actions, states and processes (13.91%), X: Psychological actions, states and processes (12.97%), and G: Government, law and order (12.97%). The prominence of abstract, social, psychological, and governmental meanings reflects the institutional and deliberative nature of parliamentary discourse. Examples include *bring + certainty/provision* in domain A, *ask + Committee/Leader* in domain S, *bring + information/knowledge/scheme/subject/system* in domain X, and *ask + Chancellor/Parliament/constituent* in domain G.

Table 45 shows that NEWS-specific collocates are concentrated in domains associated with social actors, economic activity, public issues, entertainment, and mental or informational content. The five largest semantic domains are S: Social actions, states and processes (18.29%), I: Money and commerce in industry (10.30%), A: General and abstract terms (9.77%), K: Entertainment, sports and games (8.88%), and X: Psychological actions, states and processes (8.70%), together accounting for 55.74% of the NEWS-specific set. Examples such as *ask + adviser/assistant/boss/boy/chief/daughter/family* in domain S relate to social roles and relationships, while *ask + agent/bank/customer/employee/investor/price* in domain I reflect economic and commercial contexts. Other examples, such as *bring + chaos/dimension/luck/problem/quality/results/sense* in domain A, *bring + game/music/player* in domain K, and *ask + information/opinion* in domain X, further reveal the broad topical range of

newspaper discourse.

ACAD-specific collocates, presented in Table 46, are strongly associated with abstract, communicative, psychological, and social meanings. Together, the four largest semantic domains account for more than half of the ACAD-specific set: A: General and abstract terms (22.65%), Q: Language and communication (13.03%), X: Psychological actions, states and processes (12.82%), and S: Social actions, states and processes (6.62%). Compared with the other registers, ACAD displays a particularly high concentration of abstract object collocates, consistent with the more conceptual and analytical nature of academic writing. Representative examples include *bring + dispute/element/gain/variable* in domain A, *ask + interviewee/speaker* in domain Q, *find + echo/concentration* in domain X, and *find + ally/association/relationship* in domain S.

4.6 Cross-verb synthesis and summary

In summary, this chapter has examined the grammatical and collocational behaviour of 25 lexical verbs across four registers. It began with an overview of their grammatical behaviour based on automatically extracted data, followed by a detailed case study of *say* that evaluated the extracted data and examined its manually validated relational collocations. The analysis was then extended to the full set of 25 verbs, first comparing the distribution and register-distinctive patterns of relational collocations across the three selected grammatical relations, and subsequently examining the extent to which these collocations are shared across registers or specific to individual registers, including the semantic patterns of object collocations.

Collectively, these empirical findings demonstrate that relational collocations of the 25 high-frequency lexical verbs are dynamically shaped by the interaction of register, grammatical relation, and individual lexical verb. Differences occur not only in the distribution and significant overuse of particular grammatical relations, but also in the patterns of shared and register-specific relational collocations and in the semantic patterns of the collocates occurring in these relations. Thus, the influence of register on verb collocations is both verb-specific and relation-specific, and extends beyond differences in frequency and distribution to broader semantic preferences reflected in the semantic domains of relational collocations.

The following Discussion chapter interprets the observed register variation in the relational collocations of the lexical verbs in relation to the situational characteristics, communicative purposes, and discourse functions of the four registers.

Table 43

Semantic Domains of Register-Specific Object Collocates in CONV

Semantic domain	No. of collocates (%)	Representative relational collocations
F: Food and farming	80 (18.14%)	bring + bread; bring + wine; call + restaurant; find + recipe; get + drink; leave + milk; make + bread; make + breakfast
O: Substances, materials, objects and equipment	77 (17.46%)	ask + thing; bring + bottle; bring + glass; bring + stuff; find + box; find + stuff; find + switch; get + bottle
B: The body and the individual	49 (11.11%)	bring + clothes; get + bag; get + cancer; get + eye; get + hair; get + leg; get + shoe; give + hand
Q: Language and communication	47 (10.66%)	bring + book; find + article; find + card; find + phone; find + video; get + book; get + email; get + phone
S: Social actions, states and processes	36 (8.16%)	ask + dad; ask + guy; bring + friend; bring + girl; bring + girlfriend; call + aunt; call + dad; call + guy
M: Movement, location, travel and transport	23 (5.22%)	call + town; find + flight; get + bus; get + car; get + flight; get + taxi; get + train; go + place

Table 43*Semantic Domains of Register-Specific Object Collocates in CONV*

Semantic domain	No. of collocates (%)	Representative relational collocations
L: Life and living things	19 (4.31%)	find + egg; give + monkey; need + egg; put + chicken; put + egg; put + potato; put + tree; see + cat
T: Time	18 (4.08%)	find + day; get + hour; get + month; get + week; leave + hour; leave + morning; make + day; need + day
H: Architecture, housing and the home	17 (3.85%)	call + house; get + bed; get + room; know + house; leave + window; make + bed; need + house; need + room
I: Money and commerce in industry	13 (2.95%)	bring + money; call + business; give + price; go + shop; know + money; leave + shop; leave + work; make + job
N: Numbers and measurement	9 (2.04%)	get + pair; get + space; need + number; pay + percent; say + bit; see + part; see + size; take + piece
P: Education	8 (1.81%)	call + school; go + school; know + school; know + student; know + teacher; leave + college; see + student; take + exam
Y: Science and technology	8 (1.81%)	bring + computer; bring + laptop; call + website; find + password; find + website; leave + computer; need + password; use + laptop
A: General and abstract terms	7 (1.59%)	get + degree; give + crap; make + excuse; need + reason; say + fact; say + reason; see + episode
C: Arts and crafts	7 (1.59%)	bring + camera; find + picture; get + photo; put + camera; put + photo; put + picture; see + photo
X: Psychological actions, states and processes	6 (1.36%)	give + feedback; say + idea; see + programme; try + way; use + ability; use + lounge
K: Entertainment, sports and games	6 (1.36%)	find + game; know + song; put + music; put + tape; use + gym; use + holiday
E: Emotional actions, states and processes	5 (1.13%)	make + fun; make + fuss; take + mick; take + mickey; take + piss
W: World and environment	4 (0.91%)	get + light; leave + light; need + light; put + light
G: Government, law and order	1 (0.23%)	take + passport
Z: Names and grammar	1 (0.23%)	ask + lady

Table 44*Semantic Domains of Register-Specific Object Collocates in PARL*

Semantic domain	No. of collocates (%)	Representative relational collocations
A: General and abstract terms	86 (16.17%)	bring + certainty; bring + provision; call + amendment; call + division; call + initiative; call + project; find + agreement; find + compromise
S: Social actions, states and processes	74 (13.91%)	ask + Committee; ask + Leader; ask + lordship; ask + standing; bring + member; bring + power; bring + service; call + assembly
X: Psychological actions, states and processes	69 (12.97%)	bring + information; bring + knowledge; bring + scheme; bring + subject; bring + system; call + campaign; call + plan; call + review
G: Government, law and order	64 (12.03%)	ask + Chancellor; ask + Parliament; ask + constituent; ask + country; bring + Ministers; bring + criminal; bring + government; bring + justice
I: Money and commerce in industry	63 (11.84%)	ask + business; ask + taxpayer; bring + business; bring + investment; bring + job; call + Secretary; call + colleague; call + office
Q: Language and communication	45 (8.46%)	ask + Speaker; bring + complaint; bring + debate; bring + remark; bring + transparency; call + debate; call + document; call + inquiry

Table 44*Semantic Domains of Register-Specific Object Collocates in PARL*

Semantic domain	No. of collocates (%)	Representative relational collocations
M: Movement, location, travel and transport	23 (4.32%)	ask + place; call + village; find + area; find + settlement; know + border; leave + place; make + position; mean + access
N: Numbers and measurement	22 (4.14%)	bring + amount; bring + deal; find + figure; find + majority; get + bit; get + figure; get + level; get + majority
H: Architecture, housing and the home	18 (3.38%)	ask + house; bring + house; call + chair; find + bed; find + housing; leave + Chair; leave + Chamber; mean + building
Z: Names and grammar	17 (3.20%)	ask + under-secretary; call + Bill; get + bill; get + everybody; get + roll-out; leave + backstop; need + bill; put + bill
T: Time	11 (2.07%)	get + appointment; know + background; know + date; leave + minute; leave + period; make + ends; mean + day; mean + month
O: Substances, materials, objects and equipment	10 (1.88%)	call + spade; get + balance; give + comfort; give + flexibility; need + flexibility; need + mechanism; show + flexibility; use + mechanism
E: Emotional actions, states and processes	10 (1.88%)	give + concern; give + reassurance; know + concern; mean + disruption; need + confidence; show + bravery; show + complacency; show + forbearance
B: The body and the individual	7 (1.32%)	bring + health; call + drug; find + treatment; find + vaccine; need + nurse; put + cap; put + flesh
P: Education	7 (1.32%)	ask + school; find + school; leave + education; mean + education; mean + school; see + school; take + lesson
K: Entertainment, sports and games	2 (0.38%)	ask + party; give + leave
L: Life and living things	2 (0.38%)	get + virus; use + drone
W: World and environment	2 (0.38%)	leave + environment; show + world

Table 45*Semantic Domains of Register-Specific Object Collocates in NEWS*

Semantic domain	No. of collocates (%)	Representative relational collocations
S: Social actions, states and processes	103 (18.29%)	ask + God; ask + adviser; ask + assistant; ask + boss; ask + boy; ask + chief; ask + daughter; ask + family
I: Money and commerce in industry	58 (10.30%)	ask + Office; ask + agent; ask + bank; ask + customer; ask + employee; ask + investor; ask + price; ask + secretary
A: General and abstract terms	55 (9.77%)	ask + owner; bring + chaos; bring + dimension; bring + luck; bring + problem; bring + quality; bring + result; bring + sense
K: Entertainment, sports and games	50 (8.88%)	ask + player; bring + game; bring + music; bring + player; bring + save; call + band; call + game; call + play
X: Psychological actions, states and processes	49 (8.70%)	ask + expert; ask + information; ask + opinion; bring + energy; bring + idea; bring + style; bring + success; bring + touch
O: Substances, materials, objects and equipment	41 (7.28%)	bring + ball; bring + chic; bring + comfort; bring + pan; bring + product; bring + trophy; call + line; find + corner
Q: Language and communication	30 (5.33%)	ask + journalist; ask + reporter; bring + cheer; bring + story; call + hotline; find + note; give + hint; give + interview
B: The body and the individual	30 (5.33%)	ask + doctor; bring + tear; find + drug; get + back; get + foot; give + vaccine; give + welly; know + body
G: Government, law and order	25 (4.44%)	ask + general; ask + judge; ask + voter; bring + lawsuit; bring + shame; call + officer; give + penalty; give + verdict
N: Numbers and measurement	24 (4.26%)	bring + level; find + pair; find + range; find + remains; find + successor; get + dose; give + space; know + limit

Table 45*Semantic Domains of Register-Specific Object Collocates in NEWS*

Semantic domain	No. of collocates (%)	Representative relational collocations
M: Movement, location, travel and transport	21 (3.73%)	call + move; find + car; get + nod; get + run; give + ground; know + city; leave + estate; leave + ground
H: Architecture, housing and the home	14 (2.49%)	ask + guest; ask + neighbour; ask + resident; call + House; get + flat; give + flat; know + neighbour; leave + building
E: Emotional actions, states and processes	12 (2.13%)	ask + fan; bring + happiness; bring + smile; find + happiness; get + kick; leave + fan; put + smile; show + mettle
T: Time	11 (1.95%)	call + day; get + start; get + update; give + day; give + update; give + year; know + day; leave + stage
Z: Names and grammar	10 (1.78%)	bring + bill; get + someone; give + mouth-to-mouth; give + thumbs-up; mean + bill; need + spring clean; try + Botox; try + app
L: Life and living things	8 (1.42%)	bring + dog; bring + life; bring + virus; get + life; leave + dead; mean + death; mean + life; work + dog
F: Food and farming	7 (1.24%)	try + cannabis; try + cream; try + diet; try + food; try + recipe; try + wine; work + farm
W: World and environment	7 (1.24%)	bring + light; bring + weather; bring + world; give + light; mean + world; see + star; want + world
P: Education	4 (0.71%)	give + lesson; need + test; show + class; show + test
C: Arts and crafts	3 (0.53%)	know + artist; show + photograph; use + camera
Y: Science and technology	1 (0.18%)	use + website

Table 46*Semantic Domains of Register-Specific Object Collocates in ACAD*

Semantic domain	No. of collocates (%)	Representative relational collocations
A: General and abstract terms	106 (22.65%)	bring + dispute; bring + element; bring + gain; bring + variable; call + evidence; call + process; call + proof; call + type
Q: Language and communication	61 (13.03%)	ask + interviewee; ask + speaker; bring + appeal; bring + argument; bring + discussion; call + clause; call + language; find + application
X: Psychological actions, states and processes	60 (12.82%)	ask + informant; ask + subject; call + method; call + procedure; call + sensation; call + skill; find + concentration; find + echo
S: Social actions, states and processes	31 (6.62%)	ask + king; bring + coalition; bring + influence; bring + institution; call + organization; find + ally; find + association; find + relationship
O: Substances, materials, objects and equipment	29 (6.20%)	call + file; find + feature; find + frame; find + material; find + string; find + structure; give + condition; give + context
G: Government, law and order	25 (5.34%)	ask + jury; bring + offender; bring + rule; bring + state; bring + war; call + Parliament; find + defendant; give + authority
N: Numbers and measurement	23 (4.91%)	find + equation; find + increase; find + rate; find + ratio; find + sequence; leave + section; make + measurement; mean + pressure
B: The body and the individual	21 (4.49%)	bring + suit; find + antibody; find + cancer; find + cell; find + disease; find + patient; get + hearing; leave + ward
M: Movement, location, travel and transport	20 (4.27%)	bring + return; call + estate; find + vessel; know + district; know + junction; know + route; know + site; leave + region
I: Money and commerce in industry	18 (3.85%)	ask + client; ask + practitioner; find + coin; leave + employment; leave + firm; make + advance; make + allowance; make + contract

Continued on next page

Table 46*Semantic Domains of Register-Specific Object Collocates in ACAD*

Semantic domain	No. of collocates (%)	Representative relational collocations
P: Education	18 (3.85%)	ask + academics; ask + class; bring + education; bring + student; get + pupil; get + student; give + training; leave + teacher
E: Emotional actions, states and processes	12 (2.56%)	find + satisfaction; get + hit; get + pleasure; get + satisfaction; give + preference; give + relief; make + peace; put + stress
Z: Names and grammar	12 (2.56%)	find + lesion; find + specie; find + tableau; need + elaboration; pay + yasak; show + cleavage; show + heterogeneity; show + lesion
T: Time	10 (2.14%)	give + period; mean + age; mean + duration; need + introduction; need + session; need + source; see + introduction; show + stage
K: Entertainment, sports and games	6 (1.28%)	mean + score; pay + piper; show + band; show + performance; show + score; use + score
Y: Science and technology	5 (1.07%)	ask + computer; call + program; see + pixel; use + microcomputer; use + program
C: Arts and crafts	3 (0.64%)	find + mosaic; give + picture; leave + designer
F: Food and farming	2 (0.43%)	show + digestion; use + heroin
H: Architecture, housing and the home	2 (0.43%)	find + passage; see + table
L: Life and living things	2 (0.43%)	show + mortality; take + root
W: World and environment	2 (0.43%)	bring + land; leave + land

5 Discussion

5.1 Introduction

This chapter discusses the major findings in relation to the three research questions and examines how register contributes to variation in the relational collocations of the 25 lexical verbs. **Section 5.2** addresses the first research question by considering the overall form and extent of register variation, drawing primarily on the cross-verb results for grammatical relations as well as shared and register-specific collocations. The case study of *say* is also revisited for its methodological role in illustrating the analytical procedure and the need to validate automatically extracted Word Sketch data. **Section 5.3** addresses the second research question by focusing specifically on how register differences are reflected quantitatively in the frequency, distribution, and significant overuse of relational collocations across verbs and grammatical relations. **Section 5.4** then addresses the third research question by relating the observed collocational and semantic differences to the communicative purposes and discourse functions of the four registers.

5.2 Overall register variation in relational collocations

The results show that the use of verb collocations varies systematically across registers, with differences shaped by both the lexical verb and the grammatical relations involved. The initial overview of grammatical behaviour across the 25 lexical verbs reveals cross-register differences in the attested grammatical relations associated with these verbs. PARL and NEWS generally display broader grammatical coverage than CONV and ACAD, although differences are relatively limited for general clause-level relations and more evident for particle-related relations. Clear contrasts emerge from the register-distinctive patterns identified for particular grammatical relations. CONV is particularly associated with pronominal objects and pronominal subjects, as well as *wh*-word patterns, whereas NEWS shows a strong association with particle-related relations. ACAD is distinguished most clearly by infinitive objects, while PARL displays fewer register-distinctive grammatical relations overall. At the verb level, the same verb may also exhibit different register-specific grammatical preferences, with different relations significantly overused in different registers. For example, *work* shows register-distinctive overuse of pronominal relations in CONV, infinitive objects in PARL, several particle-related and complement relations in NEWS, and objects together with objects of “X out” in ACAD. Register differences are therefore reflected not only in the overall range of grammatical relations attested for the verbs, but also in the verb-specific patterns of significant overuse across registers.

Further cross-register differences emerge when the focus shifts from the broader set of attested grammatical relations to relational collocations within the three selected grammatical relations. Across the 25 verbs, complete cross-register sharing is relatively limited: only 218 relational collocations are attested in all four registers, compared with 697 shared across three registers and 1,265 shared across two. These shared categories together account for 2,180 of the 5,775 relational collocations in the dataset (37.75%), whereas 3,595 collocations (62.25%) are register-specific. Register-specific relational collocations therefore constitute the larger proportion of the dataset.

The extent of cross-register overlap is uneven across register combinations, although broadly similar tendencies emerge among relational collocations shared across two and three registers. In both cases, greater overlap is generally found in combinations involving PARL and NEWS, whereas combinations containing both CONV and ACAD include considerably fewer shared relational collocations. Among the three-register combinations, PARL–NEWS–ACAD contains the largest number of shared object and modifier collocations, whereas CONV–PARL–NEWS

contains the largest number of shared complement collocations. Among the two-register combinations, PARL–NEWS consistently shows the greatest overlap across all three grammatical relations, such as *make* + *commitment* and *take* + *comfort*. These patterns should, however, be considered alongside variation in verb frequency across registers.

Cross-register sharing also varies by grammatical relation. Object and modifier collocations are attested across all possible two- and three-register combinations, whereas complement collocations occur across a more restricted set of combinations. This partly reflects verb-specific valency differences. Variation is likewise evident at the verb level: some verbs, such as *make*, show relatively extensive cross-register overlap, whereas others, including *think*, *bring*, *come*, and *try*, show little or no complete overlap across all four registers. Taken together, the extent of cross-register sharing varies according to register combination, grammatical relation, and individual lexical verb.

The semantic analysis of object collocates adds a further dimension to this variation. Shared object collocates tend to cluster in several broad semantic domains (e.g., general and abstract terms), whereas register-specific object collocates display more differentiated semantic distributions across registers. For example, 18.4% of CONV-specific object collocates belong to the semantic domain F: food and farming, including combinations such as *bring* + *bread/wine*. This suggests that register variation is also reflected in the semantic preferences of relational collocations.

Against these broader cross-verb patterns, the case study of *say* serves an important methodological purpose by illustrating the full methodology used in this study, from the examination of automatically extracted grammatical relations to the manual validation and subsequent distributional and semantic analyses of relational collocations. The validation of Word Sketch data generated in Sketch Engine showed that automatically identified collocates did not always correspond to the grammatical relations under which they were retrieved, and that the accuracy of automatic relation assignment varied across grammatical relations and verbs. The case study also provides a more detailed illustration at the level of an individual verb, with the validated results for *say* exemplifying the broader finding that the collocational behaviour of lexical verbs varies across registers and grammatical relations.

Manual validation can also reveal register differences that are not apparent from the automatically extracted Word Sketch data. The combination *go* + *mile* provides a useful example. Although *mile* was automatically identified as an object collocate of *go* in CONV, PARL, and NEWS, examination of its concordance lines revealed that the combination was used differently across these registers. In CONV, examples such as *go twenty-six miles* and *going thirty miles an hour* express distance or speed, with *mile* functioning as part of an adjunct (see Figure 6). In PARL and NEWS, by contrast, *go* + *mile* frequently occurs in the idiomatic expression *go the extra mile*, which conveys the meaning of making additional effort (see Figures 7–8). This contrast demonstrates that the same automatically extracted word combination may instantiate distinct phraseological uses across registers.

Overall, register variation in verb collocations occurs at several levels, including the grammatical relations in which individual lexical verbs occur, the extent to which relational collocations are shared across or specific to registers, and the semantic preferences associated with the collocates occurring within these relations. The form and extent of this variation are not uniform across the 25 verbs or across grammatical relations, indicating that register variation in relational collocations is both verb-specific and relation-specific. The following section examines this register variation more closely through the frequency and distribution of relational collocations.

CONCORDANCE				British National Corpus 2014 (BNC2014, spoken part)			
COL <i>mile</i> + go • 16 1.36 per million tokens • 0.00014%				Get more space			
Details				Left context	KWIC	Right context	
1	☐	☐	Discussing Expl...	d miles if that light go out they have a a halogen light to se put in front of the reflector and that can	go	twenty-six .miles	just a little old halogen light it's just an ordinary light just a light yeah because or
2	☐	☐	Discussing Expl...	but it does take seven hours and two minutes does it? yes yeah Paris to Milan mm but it can only	go	about twenty .miles	per hour in the Alps because it's oh yeah such a steep gradient yes mm mm
3	☐	☐	Discussing Expl...	coming very quickly isn't it? yeah now but the trouble is if yeah you've got sensors and they're only	going	thirty .miles	an hour you you gotta pull out to overtake them yeah but if it's a thirty mile limit yeah
4	☐	☐	Discussing Expl...	that thing on my phone's not saying to go left no oh look at all these queues but why we have to	go	forty .mile	an hour along here? it's silly that's probably why there's queues oh there's no queues
5	☐	☐	Discussing Expl...	such a bad turn off to the airport mm when you get there that erm and we missed it and we had to	go	ten .miles	out of our way to come back oh god <unclea> didn't we laugh? so it's another twenty
6	☐	☐	Discussing Advi...	then when so so then then they had to load up all their bikes and drive all the way back so they'd	gone	a hundred .miles	across Wales then load em up I drive them back the hundred miles again ah n i
7	☐	☐	Discussing • Tr...	nplex that does look nice one five eight you see I like the sound of the place that we were going to	go	.miles	away because I like the sound of the grounds and everything everybody move along we
8	☐	☐	Discussing Expl...	it's quite hard you feel like you know it felt wrong it's being <unclea> definitely since we're now	going	thirty .miles	an hour yeah I think do a next right right? and then for your um er so Stanwell Road
9	☐	☐	Discussing Expl...	n I knew how much time I had spent but in reality we're both as good as each other yeah but I just	went	the extra .mile	eight hundred times and she just sort of went oh do you know what? I wrote my ar
10	☐	☐	Discussing Expl...	light Or maybe forty miles an hour in fourth gear I was just going to say I don't think he'd make you	go	sixty .miles	an hour down the forty miles an hour in fourth gear he said yep Down the yes
11	☐	☐	Discussing Expl...	u then? Does he not think that he owes it? basically he lives in yeah our work is side okay so he's	going	twenty-five .miles	to come and get me and twenty-five miles to drop me back yeah so it's fifty mil
12	☐	☐	Discussing Expl...	then a five a four out of five? well they listen they just ignore it it was quite hard to think because	going	the extra .mile	I agree with if you're thinking of at the same time as offering praise I would put it
13	☐	☐	Discussing Expl...	nice that was actually quite interesting because but was going ooh it'll break I said the plane could	go	twenty thousand .miles	an hour it's digital you see it's not mechanical yes <unclea> for god's
14	☐	☐	Discussing Expl...	day at the office but I th I really funny when she was going on this course to it wasn't like she was	going	.miles	away weren't it? we're only in no no she could not face that but er if
15	☐	☐	Discussing Expl...	Birthday did you hear? It went over the erm <unclea> did it? yeah it	went	.miles	it went over that <unclea> right into outer space that's amazing where can it
16	☐	☐	Discussing Expl...	g there it's I mean if you feel silly don't you yeah when someone er drive up behind you and you're	going	thirty .mile	an hour on yeah mm and there's nothing at all around yeah I tell you what I'd rather li

Figure 6: Concordance lines for *go* + *mile* in BNC2014 spoken corpus (CONV)

CONCORDANCE				English parliamentary debates (ParlaMint 2.1)			
COL <i>mile</i> + go • 169 1.51 per million tokens • 0.00015%				Get more space			
Details				Left context	KWIC	Right context	
1	☐	☐	MargarettoMar	ive?</s><s>Or is it because doctors have become so demoralised that they can see no reason to	go	the extra .mile	on behalf of their patients</s><s>The NHS is excellent for acute management of
2	☐	☐	JamesTouhig	iced and failed so many people today is the lack of a role model, support, guidance and someone	going	that extra .mile	during their early years</s><s>A year or so before I retired from the Commons ,
3	☐	☐	MatthewRidley	ve examples in Committee of operators who must currently travel 50 miles to pick up someone to	go	two .miles	and then come all the way back again, and of vehicles that return empty from airports
4	☐	☐	ChristopherHolm...	</s><s>I apologise to all Chelsea fans.</s><s>This is a serious issue.</s><s>Derby County has	gone	the extra .mile	, and Cardiff and Swansea have done great stuff on access.</s><s>Arsenal's Err
5	☐	☐	MargoAlaine	been sewn up.</s><s>I am therefore particularly grateful that he was prepared to take action and	go	the last .mile	so late in the day</s><s>I also thank the Bill team and the lawyers for executing th
6	☐	☐	MikeWeatherley	ncer.</s><s>Beside providing emotional support to victims of cancer and their families, they often	go	the extra .mile	to provide information on matters such as treatment options, local support groups
7	☐	☐	AdamAfriyie	s a great opportunity to create work forces that are up to the mark, dedicated and loyal and which	go	the extra .mile	in the good times, because people with mental health challenges are a great resc
8	☐	☐	FrancisHare	tional, in that they are not that different from any other provider; it is just that they are prepared to	go	the extra .mile	to accept the family with disability and have the professional confidence to do so.
9	☐	☐	StephenPound	that, I respect that and I am proud of that.</s><s>We will hear from him later.</s><s>Can we not	go	the extra .mile	?</s><s>Can we not finally give support and succour to the Armenian people whi
10	☐	☐	EdmundCrisp	the world many examples of diseases almost being eliminated, only to return because we did not	go	the last .mile	.</s><s>It is fantastically important that we keep up that process.</s><s>It is also i
11	☐	☐	RobertGoodwill	should also like to take this opportunity to praise the work done by my officials at HS2, who have	gone	the extra .mile	to address some of the petition issues before they even needed to reach the Con
12	☐	☐	DavidAnderson2	f practical, so that we can play our part in helping families get over this?</s><s>I also urge him to	go	the extra .mile	and ensure that all Government agencies act with the utmost compassion, sensit
13	☐	☐	AudreyEmerton	the occasion and many might describe them as unsung heroes or heroines because they always	go	the extra .mile	, not just because of the NHS constitution or their code of conduct, as important s
14	☐	☐	GeorgeHowarth	cen to people for whom this is the case—the person gets a call at 11 o'clock at night telling them to	go	10 .miles	away to do a two-hour shift in a packaging factory, the first hour's earnings from which
15	☐	☐	GrahamAllen	ities again?</s><s>The movement from Whitehall to town hall is very welcome, but then we must	go	the extra .mile	.</s><s>I am sorry that the hon. Member for Glasgow Central (Alison Thewliss),
16	☐	☐	DianaWarwick	en the talent pool from which they recruit, ensuring that their schemes are accessible and, ideally,	go	the extra .mile	to ensure that those who may face disadvantages in the labour market have the c
17	☐	☐	JamesBridges	moved or have died, or they may never have existed in the first place.</s><s>However, we have	gone	the extra .mile	to find out whether these entries actually are people living at the registered addre
18	☐	☐	IanLavery	t I know and mix with are more than ordinary people; they are absolutely fantastic individuals who	go	the extra .mile	to try to help colleagues at every opportunity.</s><s>A good friend of mine, a loci
19	☐	☐	RogerLiddle	rep Britain in the EU . The politics of this is that there will be a great deal of bitterness if they have	gone	a .mile	to help the UK and we then vote to leave.</s><s>What is more, there is a significant–20
20	☐	☐	IanPaisley	nonths.</s><s>In a recent statement, the business Minister promised that the Government would	go	the extra .mile	.</s><s>Can the Minister give me any hope or encouragement this morning at Q

Figure 7: Concordance lines for *go* + *mile* in ParlaMint 2.1 corpus (PARL)

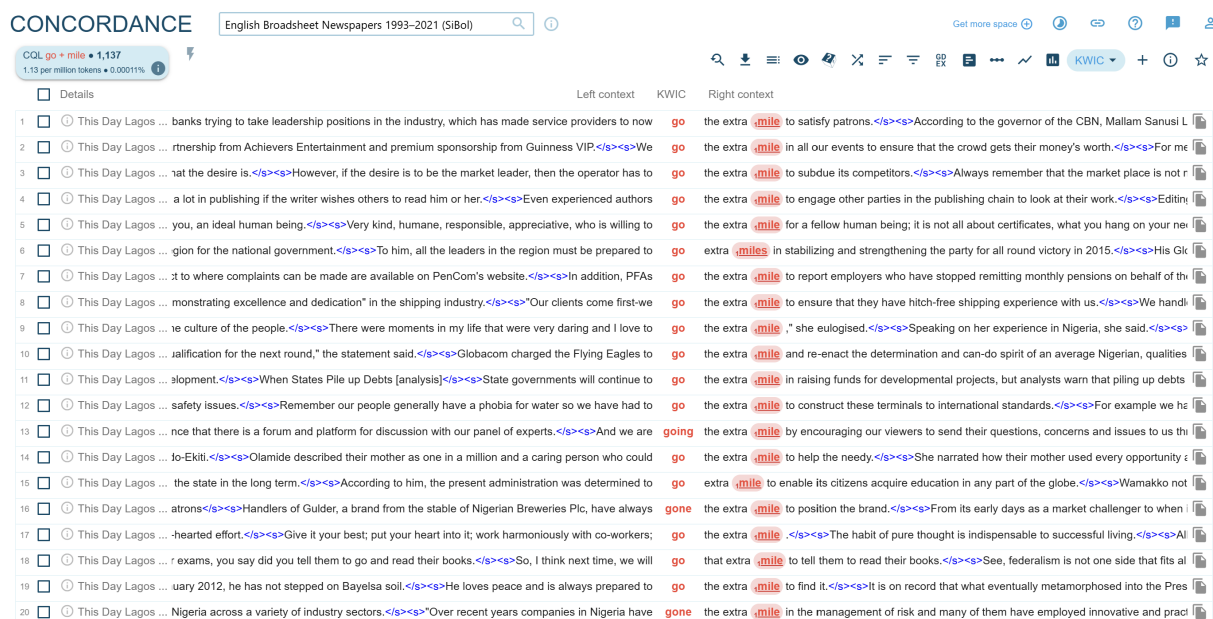


Figure 8: Concordance lines for *go + mile* in Sibol corpus (NEWS)

5.3 The influence of register on the frequency and distribution

The cross-verb analysis shows substantial register variation in the overall frequency of relational collocations. At the aggregate level, NEWS records both the highest raw frequency and the largest number of distinct relational collocates. A different pattern emerges, however, when relational collocation frequency is calculated relative to the total occurrences of the 25 verbs in each register. Relational collocations account for 23.28% of verb occurrences in PARL and 22.28% in ACAD, compared with 15.12% in NEWS and only 6.88% in CONV. The contrast between raw and proportional values highlights the need to consider cross-register differences in individual verbs frequencies when interpreting the distribution of relational collocations.

The overall distribution is broadly comparable with previous register-based phraseological research. Vilkaitė (2016), for example, found statistical collocations to be more frequent in academic prose than in newspaper language and least frequent in spoken conversation. A similar ordering is observed in the present study for ACAD, NEWS, and CONV, while PARL shows the highest proportion of relational collocations overall. This prominence of PARL is in line with Granger's (2014) finding of greater formulaicity in parliamentary debates than in editorials in English. These comparisons should be interpreted with caution, given differences in corpus design and in the operationalisation of collocation and formulaicity across studies.

The aggregate pattern is not uniform across the 25 lexical verbs. Sixteen verbs show their highest proportion values of relational collocations in PARL, whereas *go*, *know*, *come*, *look*, and *try* reach their highest values in NEWS; *give*, *think*, and *show* in ACAD; and *want* in CONV. The extent of cross-register variation also differs across verbs. For instance, relational collocations of *make* account for 23.94% of its occurrences in CONV and 51.57% in PARL, while *want* displays a different distribution, with its highest proportion in CONV. These differences indicate that the influence of register on the relative frequency of relational collocations varies across individual lexical verbs. The proportional distributions also differ across grammatical relations. Object collocations tend to reach their highest proportions in PARL for most verbs, whereas modifier collocations are distributed more widely across registers. Complement collocations are more restricted overall, but their highest proportions often occur in NEWS. Register differences in

relational collocations therefore vary according to grammatical relation as well as lexical verb. This relation-specific variation is illustrated by *leave*, for which object, modifier, and complement collocations attain their highest proportions in different registers.

The inferential results strengthen this interpretation by showing that many of the observed distributional differences are statistically significant. Significant overuse occurs in particular verb–relation cases rather than uniformly across the verb set. PARL and NEWS each contain 44 verb–relation cases of significant overuse, compared with 24 in ACAD and eight in CONV. The grammatical relations involved also differ across registers: modifier collocations show the strongest tendency towards significant overuse in PARL, whereas complement collocations are most frequently overused in NEWS. Object collocations, meanwhile, are significantly overused for many verbs in PARL, NEWS, and ACAD relative to CONV, revealing a clear contrast between CONV and the other registers. The inferential evidence thus confirms that higher relative frequencies are concentrated in particular grammatical relations for particular verbs.

Overall, register influences the frequency and distribution of verb collocations in ways that vary across lexical verbs and grammatical relations. While PARL and ACAD have the highest overall proportions of relational collocations, individual verbs do not follow this aggregate pattern consistently, and the register in which collocations are most frequent also differs by grammatical relation. Register variation in verb collocations is therefore reflected in both their relative frequency and their distribution across specific verb–relation combinations.

5.4 Collocational patterns, communicative purposes, and discourse functions

5.4.1 Introduction

The third research question considers how the collocational patterns of the 25 lexical verbs reflect the communicative purposes and discourse functions of the four registers. The register-distinctive grammatical relations identified in the preceding analyses already provide evidence. CONV, for example, is characterised by the significant overuse of pronominal relations and *wh*-word patterns, consistent with the interactive and involved nature of conversation. This accords with Biber and Conrad's (2019, pp. 118-119) observation that personal pronouns are extremely common in conversation and with Biber's (2014, p. 11) identification of “*wh*-questions, 1st and 2nd person pronouns, pronoun *it*, and indefinite pronouns” as features of involved registers.

ACAD, by contrast, is distinguished by infinitive relations, broadly consistent with Jin's (2021) findings on the use of *to*-infinitives in engineering research article discussion sections. Jin (2021) further associates *to*-infinitives, together with negation and existential *there*, with “the negation of propositions and research findings” and the identification of research gaps (p. 9). These examples illustrate how register-distinctive grammatical relations can correspond to the situational characteristics of particular registers.

Beyond these examples from CONV and ACAD, the four registers differ more broadly in their communicative orientations. As outlined in Table 3, the four registers differ in their primary communicative orientations. NEWS is primarily informational and concerned with reporting events, while ACAD is also informational but places greater emphasis on explanation and interpretation. PARL is predominantly persuasive or argumentative, with speakers debating policies and justifying positions, whereas CONV is primarily interpersonal, involving the maintenance of social relationships and the expression of attitudes and feelings. These differences in communicative purposes are reflected not only in the grammatical relations associated with the verbs, but also in the lexical choices occurring within those relations.

While grammatical relations reveal one dimension of register variation, examining the semantic patterns of collocates can further show how verb collocations reflect the communicative concerns of different registers. The following analysis therefore focuses on the semantic patterns of shared and register-specific object collocates.

5.4.2 Semantic patterns of shared object collocates

Shared object collocates cover a wide range of semantic domains, but their semantic distributions vary with the degree of register sharing. Collocates shared across all four registers are concentrated particularly in general and abstract meanings, language and communication, and psychological actions, states, and processes. By contrast, collocates shared across three or two registers show a greater concentration in domains related to social and economic meanings. This suggests that collocations shared more widely across registers tend to be semantically broader, while those shared across fewer registers are more specialised and context-specific.

The semantic profile of shared object collocates therefore reflects a broadly shared phraseological repertoire associated with the high-frequency lexical verbs. Because these collocations are not restricted to a single register, they can occur across different communicative settings and serve different discourse functions depending on context. Table 47 illustrates this pattern by presenting the relational collocations shared across the four registers.

Collocations in the domain of general and abstract terms are semantically broad and flexible. Rather than referring to a specialised topic, they serve different discourse functions across registers, ranging from interpersonal or interactional uses in CONV to institutional or deliberative functions in PARL, and evaluative roles in PARL, NEWS, and ACAD (Biber & Conrad, 2019, p. 73). In CONV, combinations such as *make + mistake*, *take + look*, *see + problem/reason*, and *take + risk* may be used to describe actions, explain problems, give reasons, and evaluate situations. In PARL, collocations such as *take + action*, *make + change/difference* can support political argumentation, deliberation, and decision-making. In NEWS, combinations such as *know + detail/fact* contribute to the reporting and explanation of public events, while other expressions help describe changes, problems, risks, and consequences. In ACAD, collocations such as *give + example* and *make + use* support explanation, exemplification, and abstract reasoning.

A second group of four-register shared collocations is concentrated in the domain of language and communication. Combinations such as *ask + question* and *give + answer* are closely related to interactivity and exchange in CONV. In PARL, collocations such as *make + statement* and *give + instruction* reflect the argumentative and procedural nature of parliamentary debates. In NEWS, communication-related collocations such as *say + word*, *give + advice*, *use + phrase*, and *put + point* contribute to reporting public statements. In ACAD, combinations such as *use + term/phrase*, *make + comment*, *show + sign*, and *make + point* support definition, argumentation, and the presentation of disciplinary knowledge.

Collocations in the domain of psychological actions, states and processes are associated with stance, cognition, attention, and evaluation. As shown in Example (8), word combinations such as *make + decision* may be used in CONV to express personal judgement, in PARL to refer to policy decisions or collective action, in NEWS to describe the decisions of public actors, and in ACAD to describe and justify research-related or methodological decisions by reference to the considerations on which they are based. Thus, although these collocations are shared across all four registers, their discourse functions are shaped by the communicative purposes of each register.

(8)	Concordance Lines of <i>Make + Decision</i> across the Four Registers
a.	a couple of days off school cos it'll only be a Thursday and a Friday and she was thinking about that wasn't she? yeah so you know cos we are just gonna go from so Wednesday night yeah to Sunday night so she'll need to make a decision on on whether cos an extra four people's a big so yeah we need to have a think cos if we have more people so the Airbnb surely it's gonna be better cos we were looking and the one that that was (<i>CONV</i>)
b.	In the course of a judicial review, the standard is not to satisfy oneself that there is a justification for the order but to satisfy oneself as to whether the execution of a discretion has been reasonable. It is sometimes described as having to demonstrate that a decision made as part of a ministerial direction is arbitrary, perverse or capricious. That is a much more limited and very different approach from that proposed in the amendments. There should be a proper warrant for something that has an enervating effect on the rights of the individual. (<i>PARL</i>)
c.	Military experts in Delhi said that if the reports were reliable—particularly the charge by one Indian military official that the mutilated body of a dead soldier was found in a forested area of the Himalayan territory—then escalation was possible unless the Indian government made a deliberate decision to “calm things down”. (<i>NEWS</i>)
d.	This decision was made based on ethical consideration despite the fact that the use of the heater probe as a rescue treatment would be a confounding variable in the assessment of the subsequent outcome (operation or death). (<i>ACAD</i>)

Table 47

Relational Collocations Shared across CONV, PARL, NEWS, and ACAD

Semantic domain	Verb collocations
A: General and abstract terms	make + change/difference/mistake/sense/use
	take + action/advantage/look/risk
	see + difference/problem/reason
	know + detail/fact
	give + detail/example/opportunity/reason
	use + form
	get + chance/hold
	work + shift
	pay + fine
	say + word
Q: Language and communication	make + arrangement/comment/point/statement
	take + note
	know + answer/name
	give + advice/answer/instruction/notice
	use + language/phrase/term/word
	find + answer/word
	put + name/point/word
	ask + question
	show + sign
	make + choice/decision/effort/progress
X: Psychological actions, states and processes	take + interest/notice/view
	give + choice/impression/information/way
	use + datum/skill
	get + idea/impression/message

Continued on next page

Table 47*Relational Collocations Shared across CONV, PARL, NEWS, and ACAD*

Semantic domain	Verb collocations
	find + solution/way
	put + effort
	pay + attention/interest
	show + interest

In addition to the three semantic domains discussed above, object collocates shared across three and two registers are more distinctively associated with social and economic meanings, suggesting that partially shared collocations fall between universally shared patterns of usage and register-specific collocational patterns.

As shown in Table 48, in the domain of social actions, states and processes, object collocates shared across three registers foreground people, family relations, social support, and rights. The CONV–PARL–NEWS and CONV–NEWS–ACAD combinations can be considered together, as both centre on everyday social experience while extending it into political discussion and news reporting. Relevant word combinations include *know + child/family/man/people/person*, *need + people/power*, and *want + help*. In comparison, PARL–NEWS–ACAD combinations, including *need + assistance/care/protection/support* and *give + power/protection/right/support*, reflect more institutional and policy-oriented perspectives, connecting political debate, public reporting, and the analysis of social issues.

Collocations shared across two registers in Table 49 display a more differentiated pattern. In the domain related to social actions, states and processes, CONV-related word combinations such as *want + person*, *see + girl*, and *ask + person* reflect interactivity. The CONV–PARL combination *say + prayer* is more specific, as it refers to a religious or ceremonial act of speaking. By contrast, PARL-related combinations foreground public, political, and institutional concerns. For example, PARL–NEWS collocations such as *make + commitment/sacrifice* and *need + guidance/leadership* are related to issues of authority, governance, and collective action, while PARL–ACAD combinations, such as *need + cooperation* and *give + assistance/consent/guidance/help/responsibility*, are related to social action. NEWS–ACAD combinations, including *ask + child/mother/panel/participant*, reflect varied social roles that can be reported in public news discourse or discussed in academic research contexts. This is illustrated in Example (9) by *ask + participant*, which appears in NEWS in reporting of journal-based findings and in ACAD in descriptions of research designs.

(9)	Concordance Lines of <i>Ask + Participant</i> in NEWS and ACAD
a.	As part of the research, published in the <i>Journal of Experimental Social Psychology</i> , American participants were asked to judge the truthfulness of trivial statements by native or non-native speakers of English, such as: “A giraffe can go without water longer than a camel can.” (NEWS)
b.	Participants were asked to rate the relevance of the questions in each dimension on a 1–5 scale in which 1 = poor, and 5 = excellent. (ACAD)

A similar tendency can be observed in the domain related to money and commerce in industry. In Table 48, CONV–PARL–NEWS collocations, such as *pay + bonus/insurance/mortgage* and *need + money/work* foreground work, money, and expenditure as practical concerns. These collo-

cations are compatible with everyday economic experience in CONV, but also enter political discussion in PARL and public reporting in NEWS. By contrast, PARL–NEWS–ACAD collocations such as *call + strike*, *find + employment*, and *pay + charge/cost/dividend/dues/employee/fare/pension* are more closely related to institutional, labour-related, and economic structures. Table 49 further sharpens this contrast, as illustrated in *pay + loan* in CONV–PARL, *make + fortune* in CONV–NEWS, *make + investment* and *need + investment* in PARL–NEWS, *bring + prosperity* in PARL–ACAD, and *show + loss* in NEWS–ACAD.

Table 48

Relational Collocations Shared across Three Registers

Register combinations	Verb collocations
<i>S: Social actions, states and processes</i>	
CONV–PARL–NEWS	know + child/family/man/people/person need + people/power ask + permission
CONV–NEWS–ACAD	make + contact get + kid leave + wife want + help bring + child
PARL–NEWS–ACAD	give + power/protection/right/support get + benefit/help/permission/support need + assistance/care/protection/service/support
<i>I: Money and commerce in industry</i>	
CONV–PARL–NEWS	need + money/work want + job/money pay + bonus/insurance/mortgage
PARL–NEWS–ACAD	call + strike pay + charge/cost/dividend/dues/employee/fare/pension find + employment

Table 49

Relational Collocations Shared across Two Registers

Register combinations	Verb collocations
<i>S: Social actions, states and processes</i>	
CONV–PARL	say + prayer want + person
CONV–NEWS	see + girl/guy/kid/man/parent
CONV–ACAD	ask + person
PARL–NEWS	make + commitment/sacrifice need + guidance/leadership want + control/independence/power/relationship/service
PARL–ACAD	give + assistance/consent/guidance/help/responsibility need + cooperation
NEWS–ACAD	work + miracle

Continued on next page

Table 49
Relational Collocations Shared across Two Registers

Register combinations	Verb collocations
	pay + homage call + conference ask + child/mother/panel/participant
<i>I: Money and commerce in industry</i>	
CONV–PARL	pay + loan
CONV–NEWS	make + fortune give + job; work + job leave + company/money pay + dollar
CONV–ACAD	make + work
PARL–NEWS	make + investment/saving need + fund/funding/investment ask + colleague/company/employer pay + penny/staff/supplier/worker
PARL–ACAD	see + role need + worker leave + market/worker pay + grant bring + prosperity
NEWS–ACAD	pay + expense/subsidy show + loss

5.4.3 Semantic patterns of register-specific object collocates

Register-specific object collocates provide a clearer view of how collocational choices are shaped by the communicative purposes and discourse functions of individual registers. CONV-specific object collocates are associated with food-related items, everyday material objects, bodily references, and interpersonal communication nouns, reflecting the role of conversation in managing ordinary activities, personal experience, and interpersonal exchange. PARL-specific object collocates are linked to abstract, social, psychological, and governmental meanings, consistent with the institutional and deliberative nature of parliamentary discourse, where speakers discuss policy issues, institutional processes, and matters of public concern. NEWS-specific object collocates are concentrated in domains of public affairs, socio-economic activity, entertainment, and informational content, reflecting the broad topic range of this register and its function of reporting and evaluating events of public relevance. ACAD-specific object collocates prioritise abstract, communicative, psychological, and social expressions, reflecting the conceptual, analytical, and knowledge-oriented nature of academic writing.

Beyond these overall semantic profiles, two register-distinctive tendencies are particularly noteworthy. Among their leading semantic domains, CONV is the only register in which the domains—F: Food and farming, B: The body and the individual appears, and M: Movement, location, travel and transport—appear, while NEWS is the only register in which the domain K: Entertainment, sports and games is prominent. These four domains point to areas of semantic specificity that are closely aligned with the communicative purposes of the two registers. The CONV-specific prominence of bodily and individual-related collocates, such as *get +*

bus/car/flight/taxi/train, bring + bread/wine, bring + clothes, get + bag/cancer/eye/ hair/leg/shoe, and give + hand, reflects the concern of conversation with physical states, everyday actions, and personal experiences. By contrast, the NEWS-specific prominence of entertainment, sports, and games aligns with the broad topical scope of news discourse, which focuses on reporting current and noteworthy events across a wide range of domains.

At the same time, register-specificity is not limited to domains that occur in only one register. Even when all four registers share the same leading semantic domains, the specific collocates classified under each domain vary in ways that reflect register-specific situational characteristics. Table 50 below illustrates this point by presenting seven semantic domains that are shared by more than one register but realised through different relational collocations.

In O: Substances, materials, objects and equipment, CONV collocates, such as *bring + bottle/glass/stuff* and *find + box/stuff/switch*, refer to concrete objects involved in everyday activities, whereas NEWS collocates, such as *bring + ball/product/trophy*, are more closely related to the reporting of sports and public consumption. ACAD collocates, such as *find + feature/frame/material/structure* and *give + condition/context*, are more abstract and analytical, reflecting the explanatory orientation of academic writing.

A similar contrast appears in Q: Language and communication. CONV includes everyday communicative objects such as *find + phone/video* and *get + email/phone*, while PARL contains collocates such as *bring + debate/remark/transparency* and *call + inquiry*, which are associated with public discussion and procedural debate. In ACAD, collocates such as *ask + interviewee/speaker, bring + argument/discussion*, and *call + clause/language* are more closely linked to research procedures, argumentation, and linguistic description.

The same pattern is also found in the domain related to social, psychological, governmental, economic, and abstract meanings. In S: Social actions, states and processes, CONV collocates mainly refer to familiar social relations, such as *ask + dad/guy* and *bring + friend/girl/girlfriend*, while PARL collocates such as *call + assembly* and *ask + Committee/Leader/lordship* reflect institutional roles. NEWS contains a wider range of publicly reportable social actors, such as *ask + adviser/assistant/boss/chief/daughter/family*, whereas ACAD includes collocates referring to social organisations and affiliations, such as *bring + institution, call + organization*, and *find + association/ally*.

In X: Psychological actions, states and processes, PARL collocates, such as *bring + information/knowledge/scheme/system*, relate to policy discussion and institutional reasoning, while NEWS collocates, such as *ask + expert/opinion* and *bring + idea/success*, support public reporting and evaluation. ACAD collocates, such as *ask + informant/subject, call + method/procedure*, and *find + concentration*, are more strongly associated with research design and analytical categorisation.

In G: Government, law and order, collocations in PARL, such as *ask + Chancellor/Parliament/constituent/country* and *bring + Ministers/government/justice*, reflect the institutional setting of parliamentary debate. In ACAD, however, collocations such as *bring + offender/rule/state/war* and *give + authority* are less tied to immediate political debate and more closely associated with legal and historical discussion.

In I: Money and commerce in industry, PARL and NEWS also differ in their lexical realisations. PARL collocations such as *ask + business/taxpayer, bring + business/investment/job*, and *call + colleague/office* connect economic issues with policy debate and public responsibility. NEWS collocations such as *ask + agent/bank/customer/employee/investor/price/secretary* are more closely related to reported economic actors, market activities, and employment relations. This suggests that economic collocates in PARL are oriented towards deliberation and governance, whereas those in NEWS are oriented towards reporting and explaining socio-economic events.

In A: General and abstract terms, the collocates show different degrees of abstraction across registers. PARL collocations such as *bring + certainty/provision*, *call + amendment/division/initiative/project*, and *find + agreement/compromise* are closely related to parliamentary procedure, negotiation, and institutional decision-making. NEWS collocations such as *bring + chaos/problem/result/sense/luck/quality* are more event-oriented and evaluative, helping report developments and assess their consequences. By contrast, ACAD collocations such as *bring + element/variable* and *call + evidence/process/proof/type* are more conceptual and analytical.

Table 50

Register-Specific Object Collocates in Selected Semantic Domains across Registers

Semantic domain	Register; n (%)	Representative relational collocates
<i>O: Substances, materials, objects and equipment</i>		
	CONV; 77 (17.46%)	ask + thing; bring + bottle; bring + glass; bring + stuff; find + box; find + stuff; find + switch; get + bottle
	NEWS; 41 (7.28%)	bring + ball; bring + chic; bring + comfort; bring + pan; bring + product; bring + trophy; call + line; find + corner
	ACAD; 29 (6.20%)	call + file; find + feature; find + frame; find + material; find + string; find + structure; give + condition; give + context
<i>Q: Language and communication</i>		
	CONV; 47 (10.66%)	bring + book; find + article; find + card; find + phone; find + video; get + book; get + email; get + phone
	PARL; 45 (8.46%)	ask + Speaker; bring + complaint; bring + debate; bring + remark; bring + transparency; call + debate; call + document; call + inquiry
	ACAD; 61 (13.03%)	ask + interviewee; ask + speaker; bring + appeal; bring + argument; bring + discussion; call + clause; call + language; find + application
<i>S: Social actions, states and processes</i>		
	CONV; 36 (8.16%)	ask + dad; ask + guy; bring + friend; bring + girl; bring + girlfriend; call + aunt; call + dad; call + guy
	PARL; 74 (13.91%)	ask + Committee; ask + Leader; ask + lordship; ask + standing; bring + member; bring + power; bring + service; call + assembly
	NEWS; 103 (18.29%)	ask + God; ask + adviser; ask + assistant; ask + boss; ask + boy; ask + chief; ask + daughter; ask + family
	ACAD; 31 (6.62%)	ask + king; bring + coalition; bring + influence; bring + institution; call + organization; find + ally; find + association; find + relationship
<i>X: Psychological actions, states and processes</i>		
	PARL; 69 (12.97%)	bring + information; bring + knowledge; bring + scheme; bring + subject; bring + system; call + campaign; call + plan; call + review
	NEWS; 49 (8.70%)	ask + expert; ask + information; ask + opinion; bring + energy; bring + idea; bring + style; bring + success; bring + touch
	ACAD; 60 (12.82%)	ask + informant; ask + subject; call + method; call + procedure; call + sensation; call + skill; find + concentration; find + echo

Continued on next page

Table 50*Register-Specific Object Collocates in Selected Semantic Domains across Registers*

Semantic domain	Register; <i>n</i> (%)	Representative relational collocates
<i>G: Government, law and order</i>		
	PARL; 64 (12.03%)	ask + Chancellor; ask + Parliament; ask + constituent; ask + country; bring + Ministers; bring + criminal; bring + government; bring + justice
	ACAD; 25 (5.34%)	ask + jury; bring + offender; bring + rule; bring + state; bring + war; call + Parliament; find + defendant; give + authority
<i>I: Money and commerce in industry</i>		
	PARL; 63 (11.84%)	ask + business; ask + taxpayer; bring + business; bring + investment; bring + job; call + Secretary; call + colleague; call + office
	NEWS; 58 (10.30%)	ask + Office; ask + agent; ask + bank; ask + customer; ask + employee; ask + investor; ask + price; ask + secretary
<i>A: General and abstract terms</i>		
	PARL; 86 (16.17%)	bring + certainty; bring + provision; call + amendment; call + division; call + initiative; call + project; find + agreement; find + compromise
	NEWS; 55 (9.77%)	ask + owner; bring + chaos; bring + dimension; bring + luck; bring + problem; bring + quality; bring + result; bring + sense
	ACAD; 106 (22.65%)	bring + dispute; bring + element; bring + gain; bring + variable; call + evidence; call + process; call + proof; call + type

Note. *n* = number of collocates. Percentages indicate the proportion of collocates in the semantic domain within the relevant register.

5.4.4 Summary of semantic patterns across registers

In interpreting how relational collocations reflect the communicative purposes and discourse functions of the four registers, two levels of evidence are considered: the semantic domains represented by the collocates and the specific lexical choices within those domains. At the broadest level, collocations shared across all four registers tend to occur in relatively broad semantic domains, including general and abstract terms, language and communication, and psychological actions, states and processes. Even within these widely shared domains, however, the same word combinations can fulfil different discourse functions depending on the register in which they occur.

A more differentiated pattern emerges for collocations shared across two or three registers. They are particularly concentrated in social and economic domains. Within these domains, the specific collocates vary across register combinations in ways that reflect differences in their communicative concerns. Combinations involving CONV tend to foreground everyday social relations and practical concerns, whereas PARL–NEWS–ACAD combinations are more closely associated with institutional, policy-related, labour, and social-support concerns. The semantic domains and the lexical choices within them therefore together reflect the different communicative concerns of the registers involved.

The clearest register differentiation is found in register-specific collocations. Both the semantic domains associated with each register and the specific lexical choices within those domains correspond closely to their respective situational characteristics and communicative purposes. CONV is associated with everyday experience and interpersonal interaction, PARL

with institutional and policy-related concerns, NEWS with public events and socio-economic reporting, and ACAD with abstract, analytical, and research-related meanings. Overall, the communicative purposes and discourse functions of the four registers are reflected both in the broader semantic preferences of relational collocations and in the specific lexical choices through which those preferences are realised.

5.5 Conclusion

This chapter has discussed the findings in relation to the three research questions, showing that register variation in verb collocations is both verb-specific and relation-specific, affecting the grammatical relations in which the verbs occur, frequency and distribution, and semantic preferences. The discussion further illustrates how selected grammatical and semantic patterns can be interpreted in relation to the communicative purposes and discourse functions of the four registers. The next chapter provides a general conclusion of the study, including major findings, limitations, directions and implications of this study.

6 Conclusion

This study investigated register variation in the collocational patterns of 25 high-frequency lexical verbs shared across four registers: informal conversation (CONV), parliamentary debates (PARL), news reports (NEWS), and academic prose (ACAD). It examined the extent of cross-register variation in verb collocations, how such variation is reflected in their frequency and distribution, and how the resulting patterns relate to the communicative purposes and discourse functions of the four registers. The analysis first compared the attested grammatical relations associated with the 25 verbs across registers and identified relations showing significant overuse, providing an initial account of register variation in their grammatical behaviour. A case study of *say* was then conducted to illustrate the analytical procedure and to evaluate the automatically extracted Word Sketch data. The main analysis subsequently examined 9,088 manually validated relational collocations within three grammatical relations, focusing on their frequency, distribution, and degree of cross-register sharing. Finally, 4,777 relational collocations in the object relation were analysed semantically to identify both the semantic domains represented by their collocates and the lexical choices within these domains. These patterns were then interpreted in relation to the communicative purposes and discourse functions of the four registers.

The findings show that register variation in relational collocations operates at several interconnected levels. At the grammatical level, the 25 lexical verbs differ across registers in both the range of attested grammatical relations and the particular relations showing significant overuse. At the collocational level, complete cross-register sharing is relatively limited, while register-specific relational collocations account for a larger proportion of the dataset. The degree of sharing also varies across register combinations, grammatical relations, and individual lexical verbs. Differences in frequency and distribution are similarly uneven. Although the aggregate percentage distributions of relational collocations reveal broad register differences, these patterns do not apply uniformly to all 25 verbs; instead, significant overuse is concentrated in particular grammatical relations associated with particular lexical verbs. Register variation is therefore both verb-specific and relation-specific rather than uniform across the verb set. At the semantic level, shared and register-specific object collocates differ across registers both in the semantic domains they represent and in the specific lexical items occurring within these domains. Together, these semantic and lexical differences reflect the different communicative orientations of the four registers.

Despite these findings, several limitations of the present study should be acknowledged. First, the retrieval limits of Sketch Engine may have affected the coverage of potential collocates, particularly for highly frequent or productive verbs. Second, manual validation was conducted by one person. Although explicit validation criteria were applied, decisions involving ambiguous or multifunctional candidates inevitably required a degree of judgement, and complete consistency cannot be guaranteed. When a candidate collocate included both valid and invalid occurrences, it was sometimes excluded instead of being retained with a frequency recalculated from valid occurrences only. Future research could adopt a more fine-grained validation procedure and involve additional annotators to improve the reliability of validation decisions. Third, the semantic classification of relational collocations, assisted by the USAS online tagger, may contain tagging errors, particularly for polysemous or context-dependent collocates. More extensive manual checking of semantic tags would strengthen future analyses, while an analysis at the semantic-subdomain level could reveal distinctions obscured by the broader semantic domains used in the present study.

Another aspect that was not examined systematically is variation in verb-specific valency. Although the findings suggest that variation across grammatical relations partly reflects differences

in the valency patterns of individual verbs, the role of valency was not investigated in detail. Future research could compare valency patterns across registers and examine how register-specific uses relate to more prototypical uses of individual verbs. In addition, the functional interpretation in the present study was selective. More detailed analysis of the concordance lines could clarify how comparable communicative purposes are realised through different lexical choices across registers. Future research could also develop a systematic functional classification of verb collocations and compare the distribution of the identified functions across registers, drawing on approaches used in research on lexical bundles.

In light of these findings, the study has implications for language description and pedagogy. The register-specific patterns identified for high-frequency lexical verbs suggest that their description should extend beyond general meanings to include their preferred grammatical relations, relational collocations, and semantic preferences associated with different communicative contexts. From a pedagogical perspective, such register-sensitive descriptions are particularly relevant to English for General Purposes (EGP) and English for Academic Purposes (EAP), where learners need to develop not only knowledge of frequent verbs but also an awareness of how their use varies across registers.

AI Statement

This dissertation has made use of ChatGPT (version 5.6) for language editing and stylistic refinement. All AI-assisted outputs have been critically reviewed and revised by the author to ensure accuracy and academic integrity.

References

- Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247. <https://doi.org/10.1016/j.jeap.2013.08.002>
- Alqarni, A. S. (2019). *Applying corpus pattern analysis to learner corpora: Investigating the pedagogical potential of the pattern dictionary of English verbs* [Doctoral dissertation, University of Birmingham]. <https://etheses.bham.ac.uk/id/eprint/9300/>
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173–195. <https://doi.org/10.1093/applin/22.2.173>
- Ansarifar, A., Shahriari, H., Staples, S., & Ghazanfari, M. (2025). A multi-dimensional analysis of thesis abstracts: Variation across academic levels and L1 backgrounds. *Journal of English for Academic Purposes*, 73, 101465. <https://doi.org/10.1016/j.jeap.2024.101465>
- Baisa, V., Maarouf, I. E., Rychlý, P., & Rambousek, A. (2015, November 25). Software and data for corpus pattern analysis. *RASLAN*. <https://www.semanticscholar.org/paper/Software-and-Data-for-Corpus-Pattern-Analysis-Baisa-Maarouf/6c77ea58e1c209542ab4737f9b0122fb>
- Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, 28–41. <https://doi.org/10.1016/j.jslw.2014.09.004>
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511621024>
- Biber, D. (2004). *Conversation text types: A multi-dimensional analysis*. <https://www.semanticscholar.org/paper/Conversation-text-types-%3A-A-multi-dimensional-Biber/ae685b82ea5f94bd77869e7fe6f88151c5d6383d>

- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8(1), 9–37.
- Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, 14(1), 7–34. <https://doi.org/10.1075/lic.14.1.02bib>
- Biber, D. (2019). Text-linguistic approaches to register variation. *Register Studies*, 1(1), 42–75. <https://doi.org/10.1075/rs.18007.bib>
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511814358>
- Biber, D., & Conrad, S. (2019). *Register, genre, and style*. Cambridge University Press.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . . : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Finegan, E., Conrad, S., & Leech, G. N. (2021). *Grammar of spoken and written English*. John Benjamins. <https://doi.org/10.1075/z.232>
- Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Brezina, V. (2018). Collocation graphs and networks: Selected applications. In P. Cantos-Gómez & M. Almela-Sánchez (Eds.), *Lexical collocation analysis: Advances and applications* (pp. 59–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-92582-0_4
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45–61. <https://doi.org/10.1017/S0267190512000074>
- Conrad, S. (2015). Register variation. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 309–329). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.018>

- Conrad, S., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*. <https://doi.org/10.1515/9783484604674.56>
- Conrad, S., & Biber, D. (2009). *Real grammar: A corpus-based approach to English*. Pearson/Longman.
- Cowie, A. P. (1994). Phraseology. In *The encyclopedia of language and linguistics* (pp. 3168–3171). Oxford University Press.
- Crossley, S. A., & Louwerse, M. M. (2007). Multi-dimensional register classification using bigrams. *International Journal of Corpus Linguistics*, 12(4), 453–478. <https://doi.org/10.1075/ijcl.12.4.02cro>
- Dupont, M., & Zufferey, S. (2017). Methodological issues in the use of directional parallel corpora: A case study of English and French concessive connectives. *International Journal of Corpus Linguistics*, 22(2), 270–297. <https://doi.org/10.1075/ijcl.22.2.05dup>
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? <http://hdl.handle.net/11693/22742>
- Ellis, N. C., Simpson-Vlach, R., Römer, U., O'Donnell, M. B., & Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 357–378). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.016>
- Erjavec, T., Ogrodniczuk, M., Osenova, P., Ljubešić, N., Simov, K., Grigorova, V., Rudolf, M., Pančur, A., Kopp, M., Barkarson, S., Steingrímsson, S., van der Pol, H., Depoorter, G., de Does, J., Jongejan, B., Haltrup Hansen, D., Navarretta, C., Calzada Pérez, M., de Macedo, L. D., . . . Rayson, P. (2021). *Multilingual comparable corpora of parliamentary debates ParlaMint 2.1*. <https://www.clarin.si/repository/xmlui/handle/11356/1432>
- Evert, S. (2005). *The statistics of word cooccurrences: Word pairs and collocations*. <http://nbn-resolving.de/urn:nbn:de:bsz:93-opus-23714>
- Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics* (pp. 1212–1248). Mouton de Gruyter. <https://doi.org/10.1515/9783110213881.2.1212>
- Forsberg Lundell, F. (2021). Formulaicity. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora*. Routledge.

- Frankenberg-Garcia, A., Lew, R., Roberts, J. C., Rees, G. P., & Sharma, N. (2019). Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL*, 31(1), 23–39. <https://doi.org/10.1017/S0958344018000150>
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155–179. <https://doi.org/10.1111/lang.12225>
- Gaspari, F. (2013). A phraseological comparison of international news agency reports published online: Lexical bundles in the English-language output of ANSA, Adnkronos, Reuters and UPI. *Corpus Linguistics and Variation in English: Focus on Non-Native Englishes*. <https://varieng.helsinki.fi/series/volumes/13/gaspari/>
- Gholaminejad, R., & Sarab, M. R. A. (2020). Academic vocabulary and collocations used in language teaching and applied linguistics textbooks: A corpus-based approach. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 26(1), 82–107. <https://doi.org/10.1075/term.00043.gho>
- Goulart, L., Gray, B., Staples, S., Black, A., Shelton, A., Biber, D., Egbert, J., & Wizner, S. (2020). Linguistic perspectives on register. *Annual Review of Linguistics*, 6, 435–455. <https://doi.org/10.1146/annurev-linguistics-011718-012644>
- Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast*, 14(1), 58–72. <https://doi.org/10.1075/lic.14.1.04gra>
- Granger, S. (2018). Formulaic sequences in learner corpora: Collocations and lexical bundles. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language* (pp. 228–247). Routledge.
- Granger, S. (2021a). Phraseology, corpora and L2 research. In *Perspectives on the L2 phrasicon* (pp. 3–22). Multilingual Matters. <https://doi.org/10.21832/9781788924863-002>
- Granger, S. (2021b). *Perspectives on the L2 phrasicon: The view from learner corpora*. Multilingual Matters.
- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, 52(3), 229–252. <https://doi.org/10.1515/iral-2014-0011>

- Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In *Phraseology* (pp. 27–49). John Benjamins Publishing Company. <https://doi.org/10.1075/z.139.07gra>
- Granger, S., & Paquot, M. (2009). Lexical verbs in academic discourse: A corpus-driven study of learner use. In *Academic writing: At the interface of corpus and discourse* (pp. 193–214). <https://doi.org/10.5040/9781474211703.ch-010>
- Granger, S., & Paquot, M. (2010). Customising a general EAP dictionary to meet learner needs. In S. Granger & M. Paquot (Eds.), *eLexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009*. Presses universitaires de Louvain.
- Granger, S., & Paquot, M. (2015). Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid. *Lexicographica: International Annual for Lexicography*, 31, 118–141.
- Gray, B., & Biber, D. (2015). Phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 125–145). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.008>
- Green, C. (2019). Enriching the academic wordlist and Secondary Vocabulary Lists with lexicogrammar: Toward a pattern grammar of academic vocabulary. *System*, 87, 102158. <https://doi.org/10.1016/j.system.2019.102158>
- Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In *Phraseology* (pp. 3–25). John Benjamins Publishing Company. <https://doi.org/10.1075/z.139.06gri>
- Gries, S. T. (2022). What do (some of) our association measures measure (most)? Association? *Journal of Second Language Studies*, 5(1), 1–33. <https://doi.org/10.1075/jsls.21028.gri>
- Gries, S. Th., Newman, J., & Shaoul, C. (2011). N-grams and the clustering of registers. *Empirical Language Research Journal*, 5(11).
- Han, S., & Shin, J.-A. (2015). Register variation in the lexico-grammatical and functional usages of MAKE and TAKE. *Korean Journal of English Language and Linguistics*, 15, 917–939. <https://doi.org/10.15738/kjell.15.4.201512.919>
- Hanks, P. (2004). Corpus pattern analysis.
- Hanks, P. (2012). *How people use words to make meanings: Semantic types meet valencies*. <https://www.semanticscholar.org/paper/How-people-use-words-to-make-meanings-%>

- Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. MIT Press.
- Hardie, A. (2014). *Log Ratio – an informal introduction – CASS*. <https://cass.lancs.ac.uk/log-ratio-an-informal-introduction/>
- Herbst, T. (2004). *A valency dictionary of English: A corpus-based analysis of the complementation patterns of English verbs, nouns, and adjectives*. Mouton de Gruyter.
- Huang, Y., & Ren, W. (2020). A novel multidimensional analysis of writing styles of editorials from *China Daily* and *The New York Times*. *Lingua*, 235, 102781. <https://doi.org/10.1016/j.lingua.2019.102781>
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins.
- Ilie, C. (2015). Parliamentary discourse. <https://doi.org/10.1002/9781118611463/wbielsi201>
- Jablonkai, R. (2009). “In the light of”: A corpus-based analysis of lexical bundles in two EU-related registers. *Working Papers in Language Pedagogy*, 3, 1–27. <https://doi.org/10.61425/wplp.2009.03.1.27>
- Jin, B. (2021). A multi-dimensional analysis of research article discussion sections in an engineering discipline: Corpus explorations and scientists’ perceptions. *SAGE Open*, 11(4), 21582440211050401. <https://doi.org/10.1177/21582440211050401>
- Jones, S., & Sinclair, J. M. (1974). English lexical collocations. *Cahiers de Lexicologie*, 24, 15–61.
- Kilgariff, A., & Tugwell, D. (2002). Sketching words. In *Lexicography and natural language processing: A festschrift in honour of B. T. S. Atkins* (pp. 125–137).
- Kilgariff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kruger, H., van Rooy, B., & Smith, A. (2019). Register change in the British and Australian Hansard (1901–2015). *Journal of English Linguistics*, 47(3), 183–220. <https://doi.org/10.1177/0075424219857114>

- Lee, S. (2025). *Variation in the use of academic collocations among native speakers of English* (SSRN Scholarly Paper No. 5370246). Social Science Research Network. <https://doi.org/10.2139/ssrn.5370246>
- Liu, D. (2008). Linking adverbials: An across-register corpus study and its implications. *International Journal of Corpus Linguistics*, 13(4), 491–518. <https://doi.org/10.1075/ijcl.13.4.05liu>
- Louw, B. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. In C. Heffer, H. Sauntson, & G. Fox (Eds.), *Words in context: A tribute to John Sinclair on his retirement*. University of Birmingham.
- Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319–344. <https://doi.org/10.1075/ijcl.22.3.02lov>
- Ma, H., & Qian, M. (2020). The creation and evaluation of a grammar pattern list for the most frequent academic verbs. *English for Specific Purposes*, 58, 155–169. <https://doi.org/10.1016/j.esp.2020.01.002>
- McArthur, T. (1981). *Longman lexicon of contemporary English*. Longman. <https://cir.nii.ac.jp/crid/1970586434817938867>
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395>
- Oxford University Press. (n.d.). *Oxford Learner's Dictionaries*. <https://www.oxfordlearnersdictionaries.com/>
- Paquot, M. (2012). The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool. In S. Granger & M. Paquot (Eds.), *Electronic lexicography*. Oxford University Press.
- Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly*, 15(1), 29–43. <https://doi.org/10.1080/15434303.2017.1405421>

- Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, 35(1), 121–145. <https://doi.org/10.1177/0267658317694221>
- Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149. <https://doi.org/10.1017/S0267190512000098>
- Paquot, M., & Naets, H. (2025). Phraseological sophistication as a multidimensional construct: Exploring the relationship between association, register specificity and frequency. *International Journal of Learner Corpus Research*, 11, forth.
- Paquot, M., Naets, H., & Gries, S. Th. (2020). Using syntactic co-occurrences to trace phraseological complexity development in learner writing: Verb + object structures in LONGDALE. In B. Le Bruyn & M. Paquot (Eds.), *Learner corpus research meets second language acquisition* (pp. 122–147). Cambridge University Press. <https://doi.org/10.1017/9781108674577.007>
- Partington, A. (2004). ‘Utterly content in each other’s company’: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131–156. <https://doi.org/10.1075/ijcl.9.1.07par>
- Phoocharoensil, S. (2025). Exploring collocational patterns and genres: An analysis of the use of poisonous and venomous in American English. *World Journal of English Language*, 15(2), 202–202.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519–549. <https://doi.org/10.1075/ijcl.13.4.06ray>
- Rayson, P., Archer, D., & Wilson, A. (2002). *Introduction to the USAS category system* [Report]. <http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf>
- Rees, G. P. (2018). *A phraseological multi-discipline approach to vocabulary selection for English for academic purposes* [Doctoral dissertation, Universitat Pompeu Fabra]. TDX (Tesis Doctorals en Xarxa). <http://www.tdx.cat/handle/10803/543839>
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7(1), 140–162. <https://doi.org/10.1075/arcl.7.06rom>

- Rychlý, P. (2008). *A lexicographer-friendly association score*. Masarykova Univerzita. <https://www.muni.cz/en/research/publications/937193>
- Schmitt, D., Schmitt, N., & Mann, D. (2011). *Mastering the Academic Word List*. Pearson Longman.
- Seoane, E., & Biber, D. (Eds.). (2021). *Corpus-based approaches to register variation*. John Benjamins. <https://doi.org/10.1075/scl.103>
- Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning*, 65(3), 533–562. <https://doi.org/10.1111/lang.12125>
- Sketch Engine. (2022). *SiBol: Corpus of English Broadsheet Newspapers 1993–2021* [Data set]. <https://www.sketchengine.eu/sibol-corpus/>
- Sketch Engine. (n.d.). *Statistics used in Sketch Engine*. Retrieved August 11, 2026, from <https://www.sketchengine.eu/statistics-used-in-sketch-engine/>
- Spina, S. (2025). A different view of phraseological complexity: The role of density, diversity and distance. *Journal of the European Second Language Association*, 9(1). <https://doi.org/10.22599/jesla.136>
- Spina, S., & Tanganelli, E. (2012). Collocations as an index for distinguishing text genres. *Corpus*, 11.
- Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7(2), 215–244. <https://doi.org/10.1075/ijcl.7.2.04stu>
- Szudarski, P. (2023). *Collocations, corpora and language learning*. Cambridge University Press. <https://doi.org/10.1017/9781108992602>
- The BNC Consortium. (2007). *The British National Corpus, version 3 (BNC XML Edition)* [Data set]. <https://www.natcorp.ox.ac.uk/>
- Vilkaitė, L. (2016). Formulaic language is not all the same: Comparing the frequency of idiomatic phrases, collocations, lexical bundles, and phrasal verbs. *Taikomoji Kalbotyra*,

8, 28–54.

- Wu, Y. (2021). Lexical bundles in English EU parliamentary discourse—variation across interpreted, translated, and spoken registers. *Compilation & Translation Review*, 14(2), 37–86. [https://doi.org/10.29912/CTR.202109_14\(2\).0002](https://doi.org/10.29912/CTR.202109_14(2).0002)
- Xiao, R. (2015). Collocation. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 106–124). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.007>