



Collocational Predictors of Polysemous Word Disambiguation in Business and General English: A Corpus-Based Study of “Charge,” “Interest,” and “Order.”

Duc Huy Nguyen*, Hong Thanh Vi Nguyen, Ngoc Thinh Pham, and Thi Nha Linh Nguyen

Van Hien University, Vietnam

Abstract

Polysemous vocabulary presents a persistent difficulty for learners of Business English. The lexemes “charge,” “interest,” and “order” each carry several unrelated senses, and learners who transfer the everyday meaning most familiar to them often miss the sense intended in business texts. This paper therefore asks whether the surrounding words make the intended sense predictable. A total of 24,369 keyword-in-context lines from COCA were divided into a Business sub-corpus of business and contractual texts and a General sub-corpus of non-specialist language. Each instance was assigned to a sense by an automated rule-based procedure built on a closed dictionary inventory, with cues weighted by position. Pearson chi-square tests with Bonferroni correction showed that the collocate immediately to the right of the node (R1) carries considerably more disambiguating information than the one immediately to the left (L1), with Cramér’s V reaching .66 at R1 against .31 at L1, and all twelve tests significant at $p < .001$. Mutual information (MI) identifies high-information cues such as “against,” “accrues,” and “desist,” and register narrowing emerges as the central mechanism by which the specialized financial and legal senses of “charge” and “interest” become conventionalized. The analysis yields explicit position-specific rules, for example R1 “rate” signalling the financial sense of “interest,” for use in ESP teaching and in Business English lexicography.

Keywords: Business English; collocational analysis; corpus linguistics; polysemy; sense disambiguation

1 Introduction

1.1 Statement of Problem

Polysemous vocabulary is a recurrent source of communicative breakdown in professional English. Words such as “interest,” “order,” and “charge” carry semantically divergent meanings across everyday, financial, managerial, and digital domains, so the same form serves radically different functions depending on context, and misidentifying the operative sense carries tangible professional consequences, from misunderstood contractual terms to ambiguous task assignments. Because prior instruction concentrates on a narrow range of senses, EFL learners tend to default to the most familiar meaning regardless of the professional context. Research on vocabulary acquisition explains why. Nation (2001) observes that knowing a word involves a wide range of features, beginning with familiarity with its written and spoken forms and the ability to associate a meaning with them, yet this depth is rarely achieved, and Schmitt (2000) notes that vocabulary knowledge is normally partial even for native speakers. Sinclair (1991) adds that a language user has available a large number of semi-preconstructed phrases constituting single choices, even though they might appear analysable into segments. Meaning in professional discourse is therefore encoded in recurrent patterns rather than isolated items, and learners without exposure to those patterns have little basis for recovering the intended sense. Existing research nonetheless addresses polysemy at the level of sense inventories and leaves unexamined a more tractable question: whether disambiguation in authentic business discourse is systematically predictable from the words co-occurring with each sense at identifiable positions. A corpus-based study of those patterns would extend collocational theory into professional vocabulary use and provide an evidence-based foundation for teaching disambiguation as a principled strategy rather than leaving it to incidental acquisition.

1.2 Research Questions

1. Which collocates immediately preceding and following the node are statistically associated with each attested sense of “charge,” “interest,” and “order” in the Business and General sub-corpora of COCA?
2. How do the sense-specific collocational profiles of the three target lexemes reflect underlying semantic relations such as metaphorical extension, metonymic shift, and register narrowing?
3. How well do these profiles distinguish the senses of each lexeme, and does that reliability depend on whether the cue appears immediately before or after the node?

1.3 Collocation and Meaning Construction in Business English

Words are rarely interpreted in isolation, their meanings being shaped by the lexical patterns in which they occur (Cevoli et al., 2021; McEnery & Brezina, 2022). The idea goes back to Firth (1957), who argues that a word's meaning cannot be specified apart from the words that regularly surround it, and Sinclair (1991) developed it into the framework of corpus, concordance, and collocation that makes patterned co-occurrence the empirical basis for description. The tendency is especially marked in Business English, where the same patterns recur within institutional discourse (Hyland, 2008) and formulaic sequences are central to professional writing (Staples et al., 2013). Frequent co-occurrence narrows the range of possible interpretations (Cowie, 1998), a narrowing that becomes cognitively routinised through repeated exposure (Durrant, 2013; Gablasova et al., 2017), so polysemous words lose in domain-specific use the flexibility they retain in general discourse (Deignan & Love, 2019).

1.3.1 Cluster Patterns in Business English

Cluster analysis identifies high-frequency word sequences, or lexical bundles, which reflect conventionalized patterns within specialized registers and which professional communication relies on for clarity (Pathmanathan et al., 2024). Biber and Barbieri (2007) find that frequently repeated combinations become familiar features of particular registers, as with “interest rate” and “order form.” Pérez-Paredes (2010) notes that co-occurring words influence interpretation and Katsarska (2024) observes that workplace communication depends on routinized patterns to reduce ambiguity, whereas Bharadwaj et al. (2025) show that the same terms remain polysemous outside professional contexts.

1.3.2 Collocational Patterns of “Charge,” “Interest,” and “Order” across Registers

Register narrowing appears when the three lexemes are compared across registers (Cambridge Dictionary, n.d.). “Interest” collocates with “rate” and “compound” in financial texts but with psychological states in general usage as in “keen interest” and “lose interest.” “Charge” signals payment in commercial settings (“service charge”) yet denotes accusation, duty, or electricity in general language (“criminal charges,” “be in charge of”). “Order” forms commercial phrases such as “place an order” but otherwise attaches to structural concepts (“in alphabetical order,” “law and order”). Repeated exposure makes such patterns recognisable and guides learners toward the intended sense (Wood, 2015).

1.4 General English and the Interpretation of Polysemous Words

Meanings in general discourse remain broader and are not restricted by institutional context, which makes general corpora less suitable than specialized ones for describing professional language (Khamis & Ho-Abdullah, 2015; Peters & Fernández, 2013), although Bowker and Pearson (2002) note a degree of overlap between General English and ESP. Li and Joannis (2021) find that the senses of polysemous words are more varied in general language, and Jimeno-Yepes et al. (2011) attribute this to the absence of register-specific collocational constraints, so a larger proportion of instances lack a single dominant cue. The same principle nonetheless operates in both registers, as Chen (2023) argues, but the collocates carrying the disambiguating information differ. The fixed phrase “in order to” illustrates the endpoint of that divergence, the original sense of “order” having been lost through grammaticalization. Nelson (2006) finds the contrast clearly: Business English restricts meaning through routinized expressions while General English allows wider interpretation, which motivates analysing the two registers separately.

1.5 Polysemy, Sense Disambiguation, and Semantic Mechanisms

Senses that are consistently triggered by the same cues develop a strong association with them over time (Gries & Durrant, 2021; Webb et al., 2013), and three mechanisms account for how such associations become conventionalized. Register-driven conventionalization restricts a sense to an institutional context through repeated use (Hoey, 2005). Grammaticalization occurs when an item acquires a grammatical function through repeated use in fixed frames, as in “in order to” (Biber & Barbieri, 2007). Semantic prosody, developed by Louw (1993) and extended by Stubbs (2001, 1993), operates when a word repeatedly occurring with items of a particular evaluative orientation acquires that orientation itself. These mechanisms provide the interpretive framework applied to the sense profiles in Phase 4.

1.6 Corpus-Based Disambiguation Methods and ESP Applications

Corpus-based sense disambiguation rests on the distribution of collocates across sense categories, which is statistically non-random and therefore amenable to quantitative analysis

(Jimeno-Yepes et al., 2011), and offers a more reliable basis for pedagogical intervention than dictionary-based accounts (Pérez-Paredes, 2010). Building on Sinclair's (1991) concordance-based method, this supports association measures such as the chi-square test and mutual information (MI), which identify collocates functioning as reliable disambiguation cues (Church & Hanks, 1990). Hunston (2002) treats a MI score of 3.0 as the threshold separating significant attraction from chance co-occurrence. Making positionally specific cues explicit offers an alternative to incidental acquisition in ESP (Chen, 2023; Wood, 2015), particularly since learners lacking register-specific phraseology are at a systematic disadvantage (Staples et al., 2013). Instruction should therefore prioritize high-frequency lexical sequences (Biber et al., 2004) and integrate collocational disambiguation into ESP curricula and lexicographic resources (Rizvić-Eminović & Dervić, 2025).

2 Methodology

This study used a quantitative corpus-based design to examine how the senses of polysemous lexemes are signalled by their collocates. Concordancing was carried out in AntConc and the statistical analysis in R.

The data came from two sub-corpora of COCA (1990-2019), one of business and contractual texts, chosen so that the collocational patterns would reflect authentic professional usage, and one of non-specialist language, which allowed the same lexemes to be compared across registers.

3 Data Analysis and Discussion

3.1 Data Analysis

The four-phase corpus-based analysis of “charge,” “interest,” and “order” was reported in this section, with the dataset of 24,369 KWIC instances across two register sub-corpora. The Business sub-corpus of authentic English-medium business and contractual texts (13,043 instances) and the General sub-corpus of non-specialist written and spoken English (11,326 instances). In addition, concordance lines were generated in AntConc 4.3.1 (Anthony, 2024) with a window of L3 to R3 around the node, and the keyword-in-context data were exported for semantic coding.

Phase 1: Concordance Extraction and Sense Coding

The three target lexemes were entered in AntConc to produce concordance lines from L3 to R3. Sense assignment was performed by an automated rule-based procedure implemented in R, using a closed inventory derived from the Oxford Learner's Dictionaries and the Cambridge Business English Dictionary and supplemented with distinctions from cognitive-semantic accounts of polysemy (Cruse, 1986; Lakoff & Johnson, 1980; Pustejovsky, 1995). Each sense was scored by matching positionally weighted cues, awarding eight points for a phrasal match, four for a cue at L1 or R1, two at L2 or R2, and one at L3 or R3, and the highest-scoring sense was assigned; this weighting reflects the finding that adjacent collocates carry the strongest probabilistic information (Sinclair, 1991).

Each instance was also tagged for semantic mechanism and register (Barcelona, 2003; Dirven & Pörings, 2003; Lakoff & Johnson, 1980). Because assignment is deterministic, the full coding is reproducible from the deposited script. Resolution rates ranged from 74.5% for “interest” to 84.1% for “order,” both in the General sub-corpus, with an overall rate of 81.7% across 24,369 instances; these figures report the proportion of tokens carrying a cue within the

window, not accuracy against an external standard. Lines lacking any cue were labelled unresolved and retained, since that residue is itself evidence for ESP pedagogy.

Phase 2: Sense Frequency Distributions and Register Effects

Phase 2 compared the proportion of each sense across the two sub-corpora, expressed as a register ratio, the Business percentage divided by the General percentage, so that values above 1 indicate a sense more characteristic of Business. For “charge” the distribution is fairly balanced, the financial-fee and verbal-demand-payment senses predominating in both sub-corpora; the legal-accusation sense rises moderately in Business (1.40), the contractual-responsibility sense declines (0.55), and the electrical-energy and physical-attack senses occur at low rates in both. Table 1 reports the distribution.

Table 1: Sense Distribution of “charge” by Register

Sense	Business f	Business %	General f	General %	Reduction
Financial fee	751	36.12	687	35.76	1.01
Verbal demand payment	613	29.49	528	27.49	1.07
Legal accusation	147	7.07	97	5.05	1.40
Electrical energy	78	3.75	72	3.75	1.00
Contractual responsibility	62	2.98	104	5.41	0.55
Physical attack	61	2.93	60	3.12	0.94
Unresolved	367	17.65	373	19.42	0.91

Note. *f* = absolute frequency; percentages are computed within each register. Register ratio as defined above.

“Interest” shows the strongest register shift of the three. The financial-yield sense rises from 25.6% in General to 44.1% in Business, the highest ratio in the dataset at 1.72, while the psychological-attention sense falls from 19.8% to 12.7% (0.64) and the legal-stake sense declines slightly (0.84); the contractual-benefit sense is close to stable (0.88). Table 2 reports the distribution.

Table 2: Sense Distribution of “interest” by Register

Sense	Business f	Business %	General f	General %	Reduction
Financial yield	2,454	44.08	760	25.63	1.72
Legal stake	984	17.68	627	21.15	0.84
Psychological attention	704	12.65	588	19.83	0.64
Contractual benefit	385	6.92	234	7.89	0.88
Unresolved	1,040	18.68	756	25.50	0.73

Note. *f* = absolute frequency. Percentages include unresolved cases; register ratio divides Business by General.

“Order” is stable across registers. The purposive-infinitive sense realised in “in order to” accounts for about 44% in both sub-corpora (1.00), reflecting a grammaticalized device rather than a register-specific sense, and the commercial-purchase sense is likewise stable at 22.2% and 23.1%. The remaining senses occur at low base rates with ratios near 1. Table 3 reports the distribution.

Table 3: Sense Distribution of “order” by Register

Sense	Business f	Business %	General f	General %	Reduction
Purposive infinitive	2,378	44.06	2,831	43.96	1.00
Commercial purchase	1,197	22.18	1,488	23.11	0.96
Organizational arrangement	290	5.37	374	5.81	0.92
Legal directive	228	4.22	271	4.21	1.00
Militaristic command	204	3.78	242	3.76	1.01
Organizational state	192	3.56	213	3.31	1.08
Unresolved	908	16.82	1,021	15.85	1.06

Note f = absolute frequency. The high proportion of purposive uses in both registers reflects the pervasive “in order to” construction, which operates independently of register-specific lexical content.

Chi-square tests of sense by register confirm these patterns: the distributions differ significantly for “charge,” $\chi^2(5, N = 3,260) = 22.11, p < .001, V = .08$, and for “interest,” $\chi^2(3, N = 6,736) = 251.30, p < .001, V = .19$, but not for “order,” $\chi^2(5, N = 9,908) = 2.32, p = .80, V = .02$, whose sense distribution is statistically indistinguishable across registers, consistent with its grammaticalized profile.

The coefficient of variation for collocate type frequencies ranges from 233% to 1,462% (median 454%), the highest values occurring for “order,” whose R1 slot is dominated by the single form “to.” Such dispersion reflects lexical diversity rather than instability of the cues used in Phase 3.

Phase 3: Collocational Disambiguation Analysis

Phase 3 tested whether the collocational environments at L1 and R1 are non-randomly associated with sense realization. For each lexeme in each register, a contingency table set senses against the twenty most frequent collocate types at that position, with remaining collocates pooled and senses under ten instances excluded to satisfy expected-frequency assumptions. Pearson chi-square tests were computed in R with Bonferroni correction across the twelve tests, and Cramér's V was used as effect size (Cohen, 1988). All twelve tests were significant after correction ($p < .001$), with Cramér's V ranging from .176 to .655. As Table 4 shows, R1 effect sizes are consistently larger than L1 effect sizes for all three lexemes in both registers, the asymmetry indicating that the right-adjacent collocate carries more disambiguating information than the left-adjacent one.

Table 4. Chi-Square Tests of Sense by Collocate at L1 and R1

Lexeme	Register	Position	χ^2	df	N	Cramér's V
charge	Business	L1	466.77	100	1,712	.234
charge	Business	R1	2,205.61	100	1,712	.508
charge	General	L1	740.53	100	1,548	.309
charge	General	R1	1,882.74	100	1,548	.493
interest	Business	L1	444.87	60	4,527	.181
interest	Business	R1	5,710.17	60	4,527	.648
interest	General	L1	204.33	60	2,209	.176
interest	General	R1	2,845.26	60	2,209	.655
order	Business	L1	1,430.48	100	4,489	.252
order	Business	R1	6,806.50	100	4,489	.551
order	General	L1	1,363.09	100	5,419	.224
order	General	R1	8,052.74	100	5,419	.545

Note. All tests significant at $p < .001$ after Bonferroni correction. Computed in R (R Core Team, 2025).

MI was then computed for each collocate-sense pair to identify the most characteristic cues; following Hunston (2002), a value of 3.0 or higher indicates significant attraction. Table 5 reports the strongest cues in the Business sub-corpus.

Table 5: Top Disambiguating Collocates for Senses in the Business Sub-corpus

Lexeme	Sense	Collocate	Position	n	PMI
charge	Physical attack	at	R1	13	4.81
charge	Physical attack	into	R1	10	4.81
charge	Physical attack	no	L1	8	4.00
charge	Electrical energy	positive	L1	8	3.65
charge	Electrical energy	negative	L1	5	3.46
charge	Legal accusation	against	R1	32	3.54
charge	Legal accusation	is	R1	35	3.50
charge	Legal accusation	assault	L1	5	2.86
interest	Contractual benefit	at	R1	35	3.56
interest	Contractual benefit	accrues	L1	15	3.56
interest	Contractual benefit	between	R1	13	3.35
interest	Contractual benefit	bearing	R1	10	3.56
interest	Psychological attention	own	L1	13	2.39
interest	Psychological attention	special	L1	29	2.15
interest	Legal stake	controlling	L1	11	2.08
interest	Legal stake	lost	L1	12	2.20
order	Militaristic command	was	R1	26	4.46
order	Militaristic command	is	R1	43	4.30
order	Legal directive	desist	L1	8	4.13
order	Legal directive	that	R1	65	4.09
order	Legal directive	executive	L1	33	3.46
order	Organizational arrangement	chronological	L1	11	3.60
order	Organizational arrangement	by	R1	17	3.72
order	Organizational state	of	L1	49	3.80
order	Organizational state	out	R1	8	3.46

Note: MI = mutual information; n = joint occurrences in the Business sub-corpus. Collocates with a joint frequency below five were excluded. Values below 3.0 are marked with a dagger.

The strongest cues cluster in the peripheral senses. Physical-attack for “charge” and militaristic-command for “order” show cues above 4.0, while the legal-stake and psychological-attention senses of “interest” reach none above 3.0.

Phase 4: Semantic Mechanism Analysis and Rule Formulation

Phase 4 accounts for the mechanisms behind the sense variation and translates the statistical patterns into disambiguation rules for ESP. Each sense was classified as literal, metaphorical extension, metonymic transfer, register narrowing, or grammaticalization. Register narrowing dominates “charge,” producing its financial-fee, verbal-demand-payment, legal-accusation, and contractual-responsibility senses, while metaphorical extension produces only the electrical-energy sense (Lakoff & Johnson, 1980). It likewise dominates “interest,” yielding the financial-yield, legal-stake, and contractual-benefit senses from the prototypical psychological-attention sense, whereas “order” draws on all five. The rules below state the L1 and R1 cues that signal each sense at non-random rates.

Disambiguation Rules for “Charge”

“Charge” realises the financial-fee sense at above-chance rates, $\chi^2(100, N = 1,712) = 2,205.61$, $p < .001$, $V = .51$, when L1 is service, monthly, annual, late, extra, or additional and R1 is for, of, fee, or fees, as in “service charge.” It realises the verbal-demand-payment sense when L1 is an agent pronoun or modal complex such as “will charge” and R1 a direct object such as customers; the legal-accusation sense when L1 is criminal, felony, murder, drug, or fraud and R1 is against (MI = 3.54), is (MI = 3.54), or was (MI = 3.38); the contractual-responsibility sense when L1 is in, take, took, or person and R1 is of or over; the electrical-energy sense when L1 is electric, static, or take (MI = 3.59); and the physical-attack sense when L1 is cavalry, soldiers, or no (MI = 4.00) and R1 is at or into (both MI = 4.81).

Disambiguation Rules for “Interest”

“Interest” realises four main senses distinguished by their collocational environments, $\chi^2(60, N = 4,527) = 5,710.17$, $p < .001$, $V = .65$. It realises the financial-yield sense when L1 is compound, simple, accrued, fixed, or variable and R1 is rate, rates, payment, payments, or income; the legal-stake sense when L1 is controlling (MI = 2.08), majority, minority, beneficial, equity, or vested and R1 is in, of, or holders; the psychological-attention sense when L1 is special (MI = 2.13), keen, strong, or own (MI = 2.46) and R1 is in, to, or of; and the contractual-benefit sense when L1 is best, public, mutual, or conflict and R1 is at, bearing, over (all MI = 3.56), or between (MI = 3.35), as in “best interest” and “conflict of interest.” The first two senses reach no cue above 3.0 and rest on chi-square evidence and phrasal patterns.

Disambiguation Rules for “order”

“Order” realises the purposive-infinitive sense at above-chance rates, $\chi^2(100, N = 4,489) = 6,806.50$, $p < .001$, $V = .55$, when L1 is in and R1 is to, that, or for; this grammaticalized pattern accounts for over 44% of instances in both registers and is the most reliable rule identified here. It realises the commercial-purchase sense when L1 is place, placed, online, or your and R1 is form, online, number, or confirmation; the legal-directive sense when L1 is court, restraining (MI = 4.30), executive (MI = 3.35), gag, or desist (MI = 4.13) and R1 is that (MI = 4.09), of, or against; the organizational-arrangement sense when L1 is alphabetical (MI = 3.22), chronological (MI = 3.54), short (MI = 3.80), or particular (MI = 3.41) and R1 is by (MI = 3.87) or of; the militaristic-command sense when L1 is direct, follow, obey, or give and R1 is was (MI = 4.46) or is (MI = 4.30); and the organizational-state sense when L1 is of (MI = 4.49) and R1 is again, now, or out (MI = 3.46).

3.2 Discussion

Answers to Research Questions

Question 1 (sense-specific collocates): Each attested sense is associated with an identifiable set of collocates, and all twelve chi-square tests of sense by collocate were significant at $p < .001$ after Bonferroni correction, confirming that sense realisation is non-randomly conditioned by the immediate collocational environment. R1 carries the majority of diagnostic cues for all three lexemes, whereas L1 contributes a smaller set of modifier signals. Two senses, the legal-stake and psychological-attention senses of “interest,” fall below the MI threshold of 3.0 (Hunston, 2002) and rest instead on contingency-table evidence and phrasal patterns.

Question 2 (semantic mechanisms): “Charge” and “interest” both develop specialised financial and legal senses from a single root meaning through register narrowing, whereas “order” draws on all five mechanisms. This difference explains the register effect observed in Phase 2: “order” remains constant across registers because its dominant sense is grammaticalized rather than register-bound while “interest” varies most because its specialised senses are register-dependent.

Question 3 (reliability of the profiles): The profiles discriminate senses effectively at R1 and considerably less so at L1. Cramér’s V at R1 is two to three times larger than at L1 across every lexeme and register, and this consistent asymmetry is the clearest quantitative finding of the study. The coefficient of variation confirms that the dispersion of collocate frequencies reflects lexical diversity rather than instability of the sense categories.

Summary and Theoretical Integration

Several design features support the validity of the analysis: a dictionary-based sense inventory as an external standard, positional weighting following Sinclair’s (1991) probabilistic principle, exclusion of cells with fewer than ten observations, Bonferroni correction across the twelve tests, and an MI threshold of 3.0 (Hunston, 2002) read alongside the coefficient of

variation. The findings converge with several independent lines of prior work. The asymmetry favouring right-adjacent collocates supports Sinclair (1991) and Liu and Tang (2023); the diagnostic role of high-information collocates aligns with Durrant (2013), Gablasova et al. (2017), and Pérez-Paredes (2010); the dominance of register narrowing is consistent with Hoey (2005) and Nelson (2006); and the grammaticalization of the purposive-infinitive sense reproduces the account of Biber and Barbieri (2007).

4 Implication and Conclusion

4.1 Implications

Theoretically, the findings extend collocational theory by quantifying the positional structure of sense disambiguation. Sinclair's (1991) idiom principle and Hoey's (2005) theory of lexical priming both hold that meaning is determined by co-text, and this analysis specifies where that determination is strongest. The Phase 3 results show a consistent positional asymmetry, the collocate immediately following the node carrying considerably more disambiguating information than the one preceding it, which indicates that positional weighting is an empirical property of sense realisation rather than a coding convenience. The study also operationalises the observation of Gries and Durrant (2021) that register shapes the cognitive salience of sense-collocate associations, offering a reproducible framework for the quantitative study of polysemy. Practically, the results bear on curriculum design, lexicography, and ESP instruction. Because the core difficulty lies in collocational knowledge rather than vocabulary size, instruction should shift toward position-specific cues delivered as if-then rules, training learners to attend first to R1, where combinations such as "interest rate," "service charge," and "place an order" reliably signal the intended sense. Business English dictionaries and ESP materials could likewise list the diagnostic collocates and their positions alongside each definition.

4.2 Limitations and Future Research Suggestions

Five limitations qualify the findings. First, the analysis covers only three high-frequency lexemes and cannot be generalised to the wider polysemous vocabulary of Business English. Second, sense assignment by a rule-based procedure working from a fixed dictionary inventory leaves approximately 18% of cases unresolved and has not yet been verified against independent human coding. Third, restricting the analysis to L1 and R1 excludes wider multi-word patterns, and several diagnostic cues fall below the MI threshold of 3.0. Fourth, the data come from COCA rather than from a purpose-built archive of workplace texts. Finally, the study is descriptive and does not test whether the rules improve learner comprehension.

These limitations point to several directions for future research. A larger set of polysemous business lexemes would test whether the positional asymmetry and the dominance of register narrowing hold more widely. The validity of sense coding could be strengthened by combining the rule-based procedure with independent human coding and a formal reliability check, or with supervised machine-learning classifiers. The collocational analysis could be deepened by modelling wider spans and by techniques such as logistic regression that evaluate several positions simultaneously, and drawing on multiple corpora would improve generalisability. Most importantly, the pedagogical claims require classroom-based validation testing whether explicit instruction in the positional rules improves sense recognition and production accuracy among EFL and ESP learners.

4.3 Conclusion

This study asked whether the senses of “charge,” “interest,” and “order” can be reliably distinguished through their collocational environments in Business and General English, and how the resulting patterns can inform English for Specific Purposes. A four-phase analysis of 24,369 keyword-in-context lines from COCA shows that sense realisation is systematically conditioned by collocation, with all twelve chi-square tests significant after Bonferroni correction and the right-adjacent collocate emerging as the most informative predictor of sense. Register narrowing conventionalises the specialised financial and legal senses of “charge” and “interest” while the dominant purposive sense of “order” is grammaticalized rather than register-bound. The contribution is a set of positionally specific disambiguation rules for teaching, testing, and lexicography.

5 Data and Code Availability

The R script used for sense coding and statistical testing, together with the collocate and cluster frequency tables for the Business and General sub-corpora, are openly available at Nguyen et al. (2026). Only derived frequency data and analysis code are distributed; full corpus text is not redistributed, in accordance with the terms of use of the source corpus.

Acknowledgement

Van Hien University, Vietnam, has sponsored this research.

References

- Barcelona, A. (Ed.). (2003). *Metaphor and metonymy at the crossroads: A cognitive perspective* (Vol. 30). Mouton de Gruyter.
- Baker, M., Francis, G., & Tognini-Bonelli, E. (Eds.). (1993). *Text and technology: In honour of John Sinclair*. John Benjamins. <https://doi.org/10.1075/z.64>
- Bharadwaj, M., Paul, V., & G, B. (2025). From strategy to understanding: Public interpretation of business performance language and its communication gaps. *EPRA International Journal of Economic and Business Review*, 13(6). <https://doi.org/10.36713/epra22812>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- 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>
- Biel, L. (2015). Phraseological profiles of legislative genres: Complex prepositions as a special case of legal phrasemes in EU law and national law. *International Journal of Specialized Communication*, 37(3-4), 14-31. <https://doi.org/10.24989/fs.v37i3-4.1286>
- Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. Routledge.
- Cevoli, B., Watkins, C., & Rastle, K. (2021). What is semantic diversity and why does it facilitate visual word recognition? *Behavior Research Methods*, 53(1), 247-263. <https://doi.org/10.3758/s13428-020-01440-1>

- Chen, L. (2023). From corpus to classroom: Teaching semi-technical Business English vocabulary. *International Journal of English Linguistics*, 13(6), 55-64. <https://doi.org/10.5539/ijel.v13n6p55>
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22-29.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cowie, A. P. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. Clarendon Press.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- Davies, C., Porretta, V., Koleva, K., & Klepousniotou, E. (2022). Speaker-specific cues influence semantic disambiguation. *Journal of Psycholinguistic Research*, 51(4), 933-955. <https://doi.org/10.1007/s10936-022-09852-0>
- Deignan, A., & Love, R. (2019). *Using corpus methods to identify subject specific uses of polysemous words in English secondary school science materials* (Working paper). White Rose Research Online.
- Dirven, R., & Pörrings, R. (Eds.). (2002). *Metaphor and metonymy in comparison and contrast*. Mouton de Gruyter.
- Durrant, P. (2013). Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory*, 9(1), 1-38. <https://doi.org/10.1515/cllt-2013-0009>
- Gablasova, D., Brezina, V., & McEnery, A. M. (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>
- Gries, S. T., & Durrant, P. (2021). Analyzing co-occurrence data. In M. Paquot & S. T. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 141-159). Springer.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Jimeno-Yepes, A., McInnes, B. T., & Aronson, A. R. (2011). Collocation analysis for UMLS knowledge-based word sense disambiguation. *BMC Bioinformatics*, 12(Suppl. 3), S4. <https://doi.org/10.1186/1471-2105-12-S3-S4>
- Katsarska, V. (2024). Linguistic altitude: The significance of formulaic language in aviation English. *Knowledge-Based Organization*, 30(2). <https://doi.org/10.2478/kbo-2024-0065>
- Khamis, N., & Ho-Abdullah, I. (2015). Exploring word associations in academic engineering texts. *3L: The Southeast Asian Journal of English Language Studies*, 21(1), 117-131. <https://doi.org/10.17576/3L-2015-2101-11>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226470993.001.0001>
- Li, J., & Joannis, M. F. (2021). Word senses as clusters of meaning modulations: A computational model of polysemy. *Cognitive Science*, 45(5), e12955. <https://doi.org/10.1111/cogs.12955>

- Liu, D., & Tang, X. (2023). The contextualized representation of collocation. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics* (pp. 799-809). Chinese Information Processing Society of China. <https://aclanthology.org/2023.ccl-1.71>
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157-176). John Benjamins. <https://doi.org/10.1075/z.64.11lou>
- McEnery, T., & Brezina, V. (2022). *Fundamentals of corpus linguistics*. Cambridge University Press.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nelson, M. (2006). Semantic associations in Business English: A corpus-based analysis. *English for Specific Purposes*, 25(2), 217-234. <https://doi.org/10.1016/j.esp.2005.02.008>
- Nguyen, D. H., Nguyen, H. T. V., Pham, N. T., & Nguyen, T. N. L. (2026). *Dataset and R code for "Collocational predictors of polysemous word disambiguation in Business and General English: A corpus-based study of 'charge,' 'interest,' and 'order'"* (Version 1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.21976694>
- Pathmanathan, S., Paramasivam, S., Ali, A. M., & Darmi, R. H. (2024). Structural patterns of lexical bundles in business studies research articles. *International Journal of Linguistics*, 16(6), 118-134. <https://doi.org/10.5296/ijl.v16i6.22431>
- Pérez-Paredes, P. (2010). Corpus linguistics and language education in perspective: Appropriation and the possibilities scenario. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 53-73). Peter Lang.
- Peters, P., & Fernández, T. (2013). The lexical needs of ESP students in a professional field. *English for Specific Purposes*, 32(4), 236-247. <https://doi.org/10.1016/j.esp.2013.05.002>
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17(4), 409-441.
- Rizvić-Eminović, E., & Dervić, M. (2025). Collocational competence in L2 English: Origin, definitions, empirical research and pedagogical implications. *Časopis Filozofskog fakulteta Univerziteta u Zenici Spoznaja*, 1(1), 118-133. <https://doi.org/10.65368/SPOZNAJA202501058>
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214-225. <https://doi.org/10.1016/j.jeap.2013.05.002>
- Stubbs, M. (1993). British traditions in text analysis: From Firth to Sinclair. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 1-34). John Benjamins. <https://doi.org/10.1075/z.64.02stu>
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91-120. <https://doi.org/10.1111/j.1467-9922.2012.00729.x>
- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury Academic.