



Semantic Ambiguity of High-Frequency Polysemous Verbs in Business English: A Corpus-Based Investigation

Nguyen Anh Huy Pham*, Cong Lai Nguyen, and Hoang Thuy Tien Le

Van Hien University, Vietnam

Abstract

In ESP, interpreting contractual language depends on distinguishing polysemous items, yet corpus-based research on how to discriminate between high-frequency polysemous verbs in business and everyday English remains scarce. Using a closed sense inventory built from the Oxford Learner's and Cambridge dictionaries, two trained coders classified 36,219 KWIC instances of "run," "execute," "bear," and "discharge" drawn from the Business and General sub-corpora of COCA (15,841 Business; 20,378 General), reaching an inter-coder Cohen's kappa of .87. The Business set comprises business and finance texts including trade journalism, corporate reporting, and contractual prose while the General set draws proportionally from fiction, magazine, newspaper, and spoken material. Collocates were extracted across an L3 to R3 window and assessed with Pearson chi-square tests under Bonferroni correction and with pointwise mutual information. Sixteen tests returned significant results at $p < .001$, with Cramér's V ranging from .11 to .61. The contractual senses of "bear," "discharge," and "execute" arise through register narrowing, whereas the dominant business sense of "run" develops through metonymic transfer. Right-adjacent collocates carry more disambiguating information than left-adjacent ones, and the resulting positional rules are transferable to other polysemous verbs. The findings support sense-aware ESP vocabulary instruction and offer lexicographers an evidence base for ordering senses by register.

Keywords: Business English, collocational analysis, corpus linguistics, ESP, polysemy

1 Introduction

1.1 Statement of Problem

Interpreting commercial contracts and legal communication in English for Specific Purposes (ESP) requires not only grammatical proficiency but the capacity to determine which sense a polysemous verb carries (Mićović & Beko, 2022). Polysemous items are single orthographic

forms with several historically related meanings (Cruse, 1986; Lyons, 1977). In contracts, partnership agreements, and negotiations, interpretive ambiguity carries traceable risk: Grant et al. (2014) note that a contract can be implemented only if the parties agree on which conditions apply since divergent interpretations of the required actions lead to disputes and litigation. The problem is not confined to one learner group. Ranta (2022) argues that the ESL and EFL distinction originated in how English spread historically and overlooks the psycholinguistic similarity between the two groups, both acquiring English on top of a first language, so that similar processing produces similar outcomes.

1.2 Research Questions

1. What are the dominant sense distributions of “run,” “execute,” “bear,” and “discharge” in the Business and General sub-corpora of COCA, as identified through collocational analysis at positions L3 to R3?
2. What semantic mechanisms including register-driven narrowing, metonymic transfer, and metaphorical extension, account for the observed sense variation of these verbs across the two registers?

1.3 Semantics and Ambiguity in Business Discourse

The vocabulary of Business English (BE) is distinguished by recognizable lexical and semantic features, and Hismetollaevna et al. (2020) identify the abundance of words carrying multiple meanings as one defining characteristic, noting that such meanings are interpreted in relation to the specific contexts in which words combine. This motivates the selection of four high-frequency verbs whose disambiguation has practical consequences for professional communication (Zhang, 2009). Legal and economic research has documented those consequences: Grant et al. (2014) argue that varying interpretations of contract terms are a structural cause of disputes, while Mouritsen (2019) notes that ambiguity arises whenever a word can be construed in several ways and that context-dependency is inherent in contractual language. Mićović and Beko (2022) show that disambiguation causes significant difficulty for ESP students in legal and business communication, and Mămăligă (2025) frames this interpretative instability as a classroom learning problem rather than a theoretical one.

1.3.1 Polysemy and Sense Structure in the English Lexicon

Cruse (1986) argues that intentional semantic extension shapes the senses of polysemous words, which informs Phase 1 coding, and Lyons (1977) specifies the distinction between polysemy and homonymy. The Phase 4 typology of literal, metaphorical extension, metonymic transfer, register narrowing, and homonymy derives from Pustejovsky’s (1995) account of how polysemous words obtain context-specific readings by connecting semantic structure with syntactic surroundings, and Taylor (2003) contends that senses cluster around a basic meaning. Polysemy is a significant challenge for second-language learners (Nation, 2001), who make more errors when meeting such words without disambiguating context, and Nesselhauf (2005) finds convergent learner-corpus evidence that non-native speakers misuse ambiguous verbs in collocation, with errors concentrated in specialized combinations. These findings ground the study’s ESP orientation and the pedagogically framed rules of Phase 4 (Johns & Dudley-Evans, 1991; Hyland, 2019).

1.3.2. Collocation Structures in Business Discourse

Chaimongkol (2021) shows that many collocation structures combine high-frequency verbs with semantically specific nouns, the right-hand noun carrying the primary interpretive content while the verb functions delexically, and Sidupa and Wastono (2019) provide complementary evidence that right-positioned collocates define a verb’s meaning. Hismetollaevna et al. (2020)

note that noun phrases characterize BE, supporting the expectation that right-adjacent noun collocates are the primary cues for contractually relevant senses.

1.4 Semantic Mechanisms and Register Effects

Two accounts explain how polysemous verbs acquire specialized senses: cognitive-semantic extension through metaphor and metonymy, and corpus-based register conventionalization. Lakoff and Johnson (1980) establish that many conventional verb meanings arise from mappings in which an abstract target is understood through a concrete source domain, as in the contractual responsibility sense of “bear,” from carrying weight, and the temporal-extent sense of “run,” from locomotion. Barcelona (2003) develops the complementary account of metonymic transfer, in which meaning extends through contiguity within a single domain. This is visible in the managerial sense of “run,” which transfers the directional quality of locomotion to organizational operation, and in the medical sense of “discharge,” where contiguity between releasing a patient and the hospital produces a metonymic narrowing. The coding accordingly treats metonymic transfer as a distinct mechanism.

1.4.1 Register Narrowing and Sense Conventionalization

Dirven and Pörings (2003) integrate metaphorical extension and metonymic transfer as complementary pathways by which polysemy networks emerge, and Goossens (2002) shows that the two may operate together. A third mechanism restricts an extended meaning to an institutional or legal context, termed register narrowing here (Goźdz-Roszkowski, 2021), for which Sinclair (1991) provides the basis through the idiom principle, holding that language users draw on semi-preconstructed phrases functioning as single choices. Hunston (2002) provides the framework for measuring collocational association through pointwise mutual information (MI), where a value of three or higher denotes a collocation at least eight times more frequent than chance. Because MI underrepresents high-frequency and grammatical collocates, Stubbs (2001) warns that it must be used alongside frequency-based and chi-square evidence, and this study accordingly cross-validates salience across measures.

1.4.2 Contextual and Collocational Influences on Meaning Interpretation

A vocabulary item’s content develops from its stored semantic network as it engages with context. Gablasova et al. (2017) note that symmetrical contextual windows are an important source of evidence, Louwerse (2018) contends that words acquire part of their meaning from the linguistic and cognitive company they keep, and Labrador (2024) shows through Sketch Engine and Log Dice that quantitative datasets reveal systematic variation in verb usage. Snefjella and Kuperman (2016) add that lexical representation is influenced by emotional valence, arousal, and concreteness. Chaimongkol (2021) found congruent collocations outnumbering incongruent ones (64.48% versus 35.52%, 256 and 141 of 397) in high-frequency verb-noun collocations analyzed through English-Thai equivalence. Pérez-Sánchez and Rojas-Díaz (2021) note that many specialized phraseological units in commerce and economics originate in metaphor but are now so conventionalized that they function as dead metaphors.

1.5 Collocate Directionality and Disambiguation in Corpus Research

Sanosi (2025) argues that the positional dimension of collocation has received far less attention than frequency and association measures, documenting in academic English that most collocates stand in a consistent directional relationship with their node. Borucinsky and Kegalj (2019) converge from a different angle, showing that modifiers closer to the head noun carry more specific information and thus reduce ambiguity in complex nominal groups.

Both accounts predict that the slot immediately following the node carries a disproportionate share of sense-selecting information, the hypothesis this study tests at R1. The left of the node contributes more narrowly, through adverbials. Research verbs attract distinct collocate sets and hence distinct semantic prosodies (Al-Otaibi, 2022), that prosody originates in the adverb rather than the verb (Palma Gutiérrez, 2019), more than 95 percent of the leading adverbial collocates of “affect” occur as premodifiers (Alanazi, 2022), and the placement of epistemic adverbs follows lexical choice and the interaction of verb type with auxiliaries (Larsson et al., 2020). Left-adjacent adverbs such as “faithfully” and “duly” are therefore expected to serve as secondary but diagnostic cues.

1.6 ESP Vocabulary Instruction and Lexicographic Applications

Nesselhauf (2005) finds that even advanced learners struggle to produce verb-noun collocations accurately and links deviations to L1 and L2 alignment, while Nation (2001) argues that vocabulary knowledge extends beyond meaning to how words are used, including their collocations. Biber et al. (1999) show how features unique to professional registers emerge from systematic comparison of expert and general corpora. Mouritsen (2019) argues that corpus linguistics offers objective evidence for interpreters evaluating disputed contractual terms, Chen (2023) analyses L1 and R1 collocates for colligation and semantic prosody, and Mićović and Beko (2022) find that ESP learners struggle to identify intended senses, especially in extended uses.

2 Methodology

The data were drawn in June 2026 from two sub-corpora of the Corpus of Contemporary American English (COCA, 1990–2019). The Business sub-corpus comprises the business and finance categories, including trade journalism, corporate reporting, and contractual prose; the General sub-corpus draws proportionally from the fiction, magazine, newspaper, and spoken sections. COCA was selected because it provides authenticated native-speaker language with register metadata, permitting a register-controlled comparison of the four verbs. All instances returned by each query were retained rather than sampled, so the two sets represent the full yield in each register. Only tokens with a complete L3 to R3 window were retained, each was assigned to one of the two registers, and homonymous non-verbal uses were coded separately rather than discarded. The sense inventory was closed before coding began: for each verb the senses recorded in the two dictionaries named above were merged into one list, near-duplicate glosses were collapsed, each retained sense received a one-line definition with two worked examples, and no sense was added once coding started. Two coders trained on 100 lines per verb, discussing discrepancies against the definitions until the inventory stabilised, then coded independently from a shuffled sample, reaching Cohen’s kappa = .87.

3 Data Analysis and Discussion

3.1 Data Analysis

This section reports the four-phase analysis. Concordance lines were generated in AntConc (Anthony, 2024) with an L3 to R3 window; the dataset comprises 36,219 instances, 15,841 Business and 20,378 General. Dispersion was assessed with the coefficient of variation (Brezina, 2018), and inferential testing used Pearson chi-square tests of sense-register independence (Stefanowitsch, 2020), with Bonferroni correction across the sixteen tests and computation in R.

3.1.1 Phase 1: Concordance Extraction and Sense Coding

Each verb was entered as a node in AntConc to produce concordance lines from L3 to R3, and each instance was coded using the closed inventory described in Section 2, supplemented by cognitive-semantic accounts (Cruse, 1986; Lakoff & Johnson, 1980; Pustejovsky, 1995). A positionally weighted procedure gave highest weight to L1 and R1, following Sinclair's (1991) observation that adjacent collocates carry the strongest signal, and each instance was categorized by mechanism (Barcelona, 2003; Dirven & Pörrings, 2003). Resolution rates ranged from 41.2 percent for “execute” in General to 71.7 percent for “bear” in Business; these figures report the proportion of tokens carrying a diagnostic cue inside the window, not classification accuracy against an external standard. Tokens without such a cue were labelled unresolved and retained, since that residue shows where the L3 to R3 span is insufficient for ESP learners.

3.1.2 Phase 2: Sense Frequency Distributions and Register Effects

Phase 2 examined how the senses of each verb were distributed across the two registers. Senses were tallied by absolute frequency in each register and assigned a prominence ratio, where values above 1 indicate greater prominence in Business and values below 1 greater prominence in General. Table 1 reports the distribution for “bear.”

Table 1. Sense Distribution of “bear” by Register

Sense	Business f	Business %	General f	General %	Reduction
Animal noun (homonym)	759	21.6	659	18.9	1.14
Contractual responsibility	469	13.3	249	7.1	1.87
Evidential support	359	10.2	340	9.8	1.05
Pressure apply	327	9.3	201	5.8	1.61
Emotional tolerate	261	7.4	344	9.9	0.75
Physical carry	264	7.5	201	5.8	1.30
Reproductive yield	42	1.2	41	1.2	1.02
Directional proceed	38	1.1	51	1.5	0.74
Unresolved	995	28.3	1,401	40.2	0.70

Note. Percentages are calculated within register; *f* = absolute frequency. The prominence ratio divides the Business percentage by the General percentage; values above 1 indicate register-driven prominence in Business.

In the Business sub-corpus “bear” shifts towards contractual responsibility: collocates such as burden, expense, responsibility, obligation, risk, and loss account for 13.3 percent of Business instances against 7.1 percent in General (1.87) while the toleration sense falls from 9.9 to 7.4 percent (0.75). The homonymous noun sense remains the most frequent at 21.6 percent, occurring more in combinations such as bear market than in zoological contexts. The shift is therefore a redistribution among existing senses rather than a new one. Table 2 reports the distribution for “discharge.”

Table 2. Sense Distribution of “discharge” by Register

Sense	Business f	Business %	General f	General %	Reduction
Contractual release	995	26.0	1,393	17.7	1.47
Employment termination	378	9.9	867	11.0	0.90
Medical release	209	5.5	1,438	18.2	0.30
Physical emission	107	2.8	153	1.9	1.44
Firearm or electrical	40	1.0	65	0.8	1.27
Unresolved	2,095	54.8	3,968	50.3	1.09

Note. Percentages are calculated within register; *f* = absolute frequency.

“Discharge” shows the clearest register split of the four verbs. The contractual sense of release from obligation reaches a prominence ratio of 1.47 in Business while the medical sense of releasing a patient falls to 0.30, and the physical sense of releasing matter remains marginal in

both registers. The two dominant senses therefore move in opposite directions, which is the pattern register narrowing predicts.

The distributions for “execute” are fairly consistent across registers, as Table 3 shows: the juridical signing sense rises from 7.2 to 8.4 percent (1.16), the computational and official senses become more common in Business, and the managerial carry-out sense remains most frequent in both (22.4 versus 21.9 percent; 1.02). The consistency indicates that “execute” is already conventionalised in general usage.

Table 3. Sense Distribution of “execute” by Register

Sense	Business f	Business %	General f	General %	Reduction
Juridical sign	355	8.4	236	7.2	1.16
Managerial carry out	940	22.3	712	21.9	1.02
Judicial kill	225	5.3	171	5.3	1.02
Official law	167	4.0	100	3.1	1.29
Computational run	201	4.8	124	3.8	1.26
Unresolved	2,318	55.1	1,914	58.8	0.94

Note. Percentages are calculated within register; f = absolute frequency.

“Run” shows the strongest register effect. The computational sense, marked by the collocates click, windows, scripts, and code, rises from 1.2 percent in General to 5.4 percent in Business, a ratio of 4.57 and the highest in the dataset, while the physical locomotion sense falls from 15.8 to 13.3 percent (0.84). This is metonymic transfer rather than narrowing, since the source sense remains available.

Table 4. Sense Distribution of “run” by Register

Sense	Business f	Business %	General f	General %	Reduction
Managerial operate	628	14.6	851	14.8	0.99
Temporal extend	364	8.5	527	9.2	0.92
Physical locomote	571	13.3	906	15.8	0.84
Computational execute	232	5.4	68	1.2	4.57
Electoral campaign	126	2.9	195	3.4	0.86
Flow stream	116	2.7	160	2.8	0.97
Unresolved	2,260	52.6	3,043	52.9	0.99

Note. Percentages are calculated within register; f = absolute frequency.

The coefficient of variation was computed for collocate frequencies at L1, L2, R1, and R2 by verb, register, and sense. Values ranged from about 50 percent for low-frequency senses to more than 400 percent for high-frequency senses with diverse collocates, reflecting lexical diversity rather than instability, and the most frequent collocates produced stable rank ordering across all four positions.

3.1.3 Phase 3: Collocational Disambiguation Analysis

Phase 3 tested whether the L1 and R1 environments are non-randomly associated with sense. For each verb in each register, a contingency table set senses against the twenty most frequent L1 and R1 collocate types, with senses under ten instances excluded. Pearson chi-square tests were computed in R with Bonferroni correction across the sixteen tests, and Cramér’s V was interpreted against benchmarks of .10, .30, and .50 (Cohen, 1988). All sixteen remained significant after correction ($p < .001$), with Cramér’s V from .113 to .611, L1 effects generally small to medium and R1 effects medium to large. The largest were “run” at R1 in General (.611) and Business (.561), “discharge” at R1 in General (.558), and “bear” at R1 in Business (.553), showing that R1 carries more distinguishing information than L1. MI was computed for each collocate-sense pair, a higher value indicating stronger attraction (Hunston, 2002). Above the threshold of 3.00, a small set function as strong cues: for “discharge,” L1 accidental (4.61)

with the firearm sense; for “execute,” L1 faithfully (4.27) with the official-law sense and R1 scripts and code (4.64, 4.16) with the computational sense; for “run,” click and windows (3.76) with the computational sense and against (3.71) with the electoral-campaign sense; and for “bear,” on and upon (3.26). A second group falls below it: honorable (2.33) for employment termination, hospital (2.96) for medical release, responsibility (1.91) for contractual responsibility, faithfully, full, and final (1.46) for contractual release, and shall (1.04) for the juridical signing sense. These still imply associations two to four times stronger than chance, reflecting the tendency of MI to underrepresent high-frequency items (Stubbs, 2001), so their value rests on the chi-square associations and joint frequencies. The proportion of disambiguating, ambiguous, and no-cue instances varies by verb and register: “bear” has the highest proportion of disambiguating contexts in Business (57.4 percent) while “execute” has the highest ambiguous proportion (54.7 percent). A strong chi-square association therefore does not rule out ambiguity where more than one reading is plausible.

3.1.4 Phase 4: Semantic Mechanism Analysis and Rule Formulation

Phase 4 converts these patterns into disambiguation rules and identifies the mechanism behind each sense. Each sense from Phase 1 was assigned to one of five mechanism categories. In “bear,” metaphorical extension from carrying weight accounts for 1,028 Business instances (Lakoff & Johnson, 1980) while register narrowing generates the contractual responsibility sense (821 instances). “Run” extends by metonymic transfer to managerial operation and election campaign, and by metaphorical extension to temporal extent, computational execution, and flow. “Discharge” extends by metonymic transfer to medical release and by register narrowing to contractual release, and “execute” by register narrowing to the juridical signing and official-law senses and by metonymic transfer to managerial carry-out. The rules below specify the L1 and R1 cues that signal each contractually relevant sense at non-random rates. Where a MI value is given, 3.00 or above indicates strong attraction (Hunston, 2002); lower values indicate cues whose reliability rests on joint frequency and the chi-square association.

Disambiguation Rules for “bear”

Rules for “bear.” The contractual responsibility sense is realized, $\chi^2(140) = 5,394.87$, $p < .001$, $V = .55$, when R1 is a noun of obligation or harm (responsibility, burden, cost, liability, risk, loss) and L1 is a modal or infinitive marker. The pressure-apply sense appears when R1 is on, upon, or against (on, upon MI = 3.26); the evidential sense when R1 is mind, witness, or resemblance, as in “bear witness;” and the animal-noun reading when L1 is polar or grizzly or R1 is market or Stearns.

Disambiguation Rules for “discharge:” Rules for “discharge.” The contractual release sense is realized, $\chi^2(80) = 518.19$, $p < .001$, $V = .27$ when L1 is faithfully, full, or shall and R1 is a noun of duty; the adverbial cues reach only MI = 1.46, so the rule rests on the chi-square association and frequency. The employment-termination sense appears when L1 is honorable (2.33); the medical-release sense when L1 is hospital or after (2.96); the firearm sense when L1 is accidental or electrical (4.61); and the physical-emission sense when L1 is vaginal or nasal and R1 is into.

Disambiguation Rules for “execute:” Rules for “execute.” The juridical signing sense is realized, $\chi^2(64) = 227.53$, $p < .001$, $V = .17$, when R1 is a legal-instrument noun (agreement, contract, deed) and L1 is duly, properly, or shall (MI = 1.04, so the sense rests on the R1 field). The managerial carry-out sense appears when R1 is plan, strategy, or project; the official-law sense when L1 is faithfully (4.27) and R1 is laws or office; the computational sense when R1 is command, scripts, or code (4.64, 4.16); and the judicial-kill sense when R1 is him, her, or them.

Disambiguation Rules for “run:” Rules for “run.” The managerial operate sense is realized, $\chi^2(100) = 481.32$, $p < .001$, $V = .22$, when R1 is an organizational noun and L1 is an agent pronoun or modal-infinitive. The computational sense appears when L1 is click or R1 is program, script, or windows ($PMI = 3.76$); the temporal extent sense when L1 is long and R1 is for or until (2.50); the physical-locomotion sense when L1 is an agentive subject and R1 is away or after; and the electoral campaign sense when L1 is or was and R1 is for or against (against 3.71).

3.2 Discussion

Answer to Research Questions

Question 1 (dominant sense distributions): The Business register systematically reshapes polysemy: each verb acquires a dominant contractual or specialized sense, with the strongest effects for “discharge,” where contractual release rises against a sharply falling medical sense, and “run,” where the computational sense reaches the highest ratio at 4.57, while general and physical senses recede. Phase 3 shows that this reshaping operates through non-random collocational association, since all sixteen chi-square tests were significant at $p < .001$ after Bonferroni correction and Cramér’s V at R1 (.417 to .611) was consistently larger than at L1 (.113 to .274). The most reliable disambiguators are therefore right-adjacent complements, reinforced by a limited set of left-adjacent adverbial cues such as faithfully.

Question 2 (semantic mechanisms): The mechanism analysis shows several configurations: the business sense of “run” develops through metonymic transfer, whereas the contractual senses of “bear,” “discharge,” and “execute” arise through register narrowing. Register narrowing conventionalizes the contractual meanings through recurrent expressions such as “bear the costs” and “faithfully execute the office,” supported by both high-MI and lower-MI cues.

Summary and Theoretical Integration

Four kinds of evidence support the analysis. Structural validity rests on a closed sense inventory drawn from two dictionaries; internal validity on positional weighting and on the retention of unclassifiable cases; coding reliability on inter-coder agreement of Cohen’s $\kappa = .87$; and statistical validity on 36,219 coded instances, Bonferroni correction, Cramér’s V , the coefficient of variation, and cross-test comparison in R. Prominence was read against the MI threshold of 3.00 (Hunston, 2002), with chi-square and joint frequency carrying the weight for lower-MI cues (Stubbs, 2001).

The positional findings converge with prior research: R1 dominance is reported by Chaimongkol (2021), Sidupa and Wastono (2019), Sanosi (2025), and Borucinsky and Kegalj (2019), while left-adjacent adverbs are documented as diagnostic by Al-Otaibi (2022), Palma Gutiérrez (2019), and Alanazi (2022). The mechanism analysis aligns with the metaphor and metonymy accounts of Lakoff and Johnson (1980), Barcelona (2003), and Dirven and Pörings (2003), and register narrowing is consistent with Sinclair’s (1991) idiom principle and the framework of Biber et al. (1999).

4 Implications and Conclusion

4.1 Implications

Theoretically, the concentration of disambiguating information at R1 supports Sinclair’s (1991) claim that meaning is generated through semi-preconstructed phrases rather than isolated words, and the similar patterns noted by Chaimongkol (2021) and Sidupa and Wastono

(2019) suggest that collocational asymmetry is a general property of high-frequency verbs. The predominance of contractual senses in Business English gives empirical support to register-driven conventionalization (Goźdz-Roszkowski, 2021) and links it to the metonymy and metaphor accounts of Lakoff and Johnson (1980) and Barcelona (2003). Methodologically, because several contractual cues fell below the conventional MI threshold, MI alone cannot identify the most informative collocates of high-frequency verbs; reading it alongside chi-square association and raw frequency is more dependable (Stubbs, 2001). Practically, Phase 4 translates the statistical patterns into explicit if-then guidelines consistent with Nation's (2001) framework and the ESP tradition of Johns and Dudley-Evans (1991) and Hyland (2019). Learners can be trained to attend first to the R1 environment, where contractual collocates such as "bear the costs" reliably indicate the operative sense. The study also offers lexicography a systematic basis since Business English dictionaries could list the diagnostic collocates and their positions alongside each definition, and the same cues can inform vocabulary exercises, tests, and professional-English materials.

4.2 Limitations and Future Research Suggestions

Because the study examines only four verbs, the disambiguation criteria cannot yet be extended to the wider polysemous vocabulary of Business English. Wider contextual spans were not investigated, and for several senses only about one-fifth to one-third of occurrences were disambiguated within the L3 to R3 frame. Sense classification relied on a predefined inventory, which may not capture emerging or less conventional senses. The data come from COCA rather than a purpose-built collection of contractual texts, which may limit the applicability of the findings to the contractual register.

The most direct extension is to widen the collocational span and model multi-word patterns with techniques such as logistic regression. A dedicated corpus of authentic contracts would allow the rules to be tested in the register where they matter most, and extending the analysis to further verbs would test the generality of the R1-dominance pattern. Classroom intervention research could establish whether teaching the rules improves learners' sense recognition and production accuracy.

4.3 Conclusion

This study examined how "run," "execute," "bear," and "discharge" are disambiguated across Business and General English. A four-phase analysis of 36,219 KWIC instances showed that sense realization is systematically conditioned by collocation, with all sixteen chi-square tests significant at $p < .001$ after Bonferroni correction and the larger R1 effect sizes establishing the right-adjacent collocate as the most informative predictor of sense. Register narrowing conventionalizes the contractual senses of "bear," "discharge," and "execute" while the business sense of "run" develops through metonymic transfer. The contribution is a set of positionally specific disambiguation rules and a resource for teaching, testing, and lexicography.

4.3 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 and the contingency tables underlying the chi-square and Cramér's V analyses, are openly available at Pham 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

The research was aided by Van Hien University.

References

- Alanazi, Z. (2022). Corpus-based analysis of near-synonymous verbs. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), Article 15. <https://doi.org/10.1186/s40862-022-00138-5>
- Al-Otaibi, G. M. (2022). Semantic prosody of research verbs: A corpus-informed study. *Journal of Language and Education*, 8(2), 48-65. <https://doi.org/10.17323/jle.2022.12985>
- Anthony, L. (2024). *Addressing the challenges of data-driven learning through corpus tool design: In conversation with Laurence Anthony*. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 9-18). Routledge. <https://doi.org/10.4324/9781003413301-2>
- Barcelona, A. (Ed.). (2003). *Metaphor and metonymy at the crossroads: A cognitive perspective* (Vol. 30). Mouton de Gruyter.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Borucinsky, M., & Kegalj, J. (2019). Syntactic ambiguity of (complex) nominal groups in technical English. *International Journal of English Studies*, 19(2), 83-102. <https://doi.org/10.6018/ijes.352751>
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Cambridge University Press. (n.d.). *Cambridge dictionary*. Retrieved June 7, 2026, from <https://dictionary.cambridge.org/>
- Chaimongkol, T. (2021). Congruent and incongruent collocations of high-frequency verbs: Do, get, give, have, make, take. *Academic and Research Journal of Liberal Arts*, 16(1), 130-146.
- 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>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- Deignan, A., & Love, R. (2021). Using corpus methods to identify subject-specific uses of polysemous words in English secondary school science materials. *Corpora*, 16(2), 165-189. <https://doi.org/10.3366/cor.2021.0216>
- Dirven, R., & Pörings, R. (Eds.). (2003). *Metaphor and metonymy in comparison and contrast*. Mouton de Gruyter.
- 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>

- Goossens, L. (2002). Metaphonymy: The interaction of metaphor and metonymy in expressions for linguistic action. In R. Dirven & R. Pörings (Eds.), *Metaphor and metonymy in comparison and contrast* (pp. 349-377). Mouton de Gruyter.
- Goźdz-Roszkowski, S. (2021). Corpus linguistics in legal discourse. *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, 34, 1515-1540. <https://doi.org/10.1007/s11196-021-09858-1>
- Grant, S., Kline, J. J., & Quiggin, J. (2014). A matter of interpretation: Ambiguous contracts and liquidated damages. *Games and Economic Behavior*, 85(1), 180-187. <https://doi.org/10.1016/j.geb.2014.01.019>
- Hismetollaeva, B. G., Askarovna, A. S., Zylpykarovna, M. F., & Turysbekovna, O. G. (2020). Lexico-semantic and functional peculiarities of business discourse as means of communication. *Евразийский Союз Ученых*, 76(3), 4-14. <https://doi.org/10.31618/ESU.2413-9335.2020.3.76.909>
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Hyland, K. (2019). English for specific purposes: Some influences and impacts. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 337-353). Springer. https://doi.org/10.1007/978-3-319-58542-0_19-1
- Johns, A. M., & Dudley-Evans, T. (1991). English for specific purposes: International in scope, specific in purpose. *TESOL Quarterly*, 25(2), 297-314. <https://doi.org/10.2307/3587465>
- Labrador, B. (2024). A child-oriented corpus-based study of semantically related verbs using Sketch Engine. *Focus on ELT Journal*, 6(2), 50-69. <https://doi.org/10.14744/felt.6.2.4>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Larsson, T., Callies, M., Hasselgård, H., Laso, N. J., Van Vuuren, S., Verdaguer, I., & Paquot, M. (2020). Adverb placement in EFL academic writing: Going beyond syntactic transfer. *International Journal of Corpus Linguistics*, 25(2), 155-184. <https://doi.org/10.1075/ijcl.19131.lar>
- Louwerse, M. M. (2018). Knowing the meaning of a word by the linguistic and perceptual company it keeps. *Topics in Cognitive Science*, 10(3), 573-589. <https://doi.org/10.1111/tops.12349>
- Lyons, J. (1977). *Semantics: Volume 1*. Cambridge University Press.
- Mămăligă, A. (2025). *American vs. British phrasal verbs: Linguistic variation and multicultural implications*. In *Competitiveness and Innovation in the Knowledge Economy: Annual International Scientific Conference* (29th ed., pp. 300-307). Academy of Economic Studies of Moldova. <https://doi.org/10.24818/cike2025.37>
- Merriam-Webster. (n.d.). *Metaphor*. In *Merriam-Webster.com dictionary*. Retrieved June 7, 2026, from <https://www.merriam-webster.com/dictionary/metaphor>
- Mićović, D. N., & Beko, L. V. (2022). Polysemy-related problems in ESP students: A case study. *Zbornik radova Filozofskog fakulteta u Prištini*, 52(3), 123-144. <https://doi.org/10.5937/zrffp52-32382>
- Mouritsen, S. C. (2019). Contract interpretation with corpus linguistics. *Washington Law Review*, 94(3), 1337.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.14>
- Oxford University Press. (n.d.). *Oxford Learner's Dictionaries*. Retrieved June 7, 2026, from <https://www.oxfordlearnersdictionaries.com/>
- Palma Gutiérrez, M. (2019). Semantic prosody in middle construction predicates: Exploring adverb + verb collocation in middles. In G. Corpas Pastor & R. Mitkov (Eds.), *Computational and corpus-based phraseology* (pp. 345-359). Springer. https://doi.org/10.1007/978-3-030-30135-4_25
- Pérez-Sánchez, J. M., & Rojas-Díaz, J. L. (2021). "Beyond the grave": Equivalences of specialized phraseological units with dead metaphors in an English-Spanish specialized dictionary of commerce and economics. *Mutatis Mutandis: Revista Latinoamericana de Traducción*, 14(2), 527-546. <https://doi.org/10.17533/udea.mut.v14n2a12>
- Pham, N. A. H., Nguyen, C. L., & Le, H. T. T. (2026). *Dataset and R code for "Semantic ambiguity of high-frequency polysemous verbs in Business English: A corpus-based investigation"* (Version 1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.21967677>
- Pustejovsky, J. (1995). *The generative lexicon*. MIT Press. <https://doi.org/10.7551/mitpress/3225.001.0001>
- Ranta, E. (2022). From learners to users-errors, innovations, and universals. *ELT Journal*, 76(3), 311-319. <https://doi.org/10.1093/elt/ccac024>
- Rumshisky, A., & Batiukova, O. (2008). *Polysemy in verbs: Systematic relations between senses and their effect on annotation*. In *Proceedings of the Workshop on Human Judgements in Computational Linguistics (Coling 2008)* (pp. 33-41). Association for Computational Linguistics.
- Sanosi, A. B. (2025). A corpus-based analysis of collocate directionality in academic English writing. *Heliyon*, 11(2), e42088. <https://doi.org/10.1016/j.heliyon.2025.e42088>
- Sidupa, C., & Wastono, A. T. (2019). Semantic preference of verb-noun collocation: Corpus-based analysis. In *Proceedings of the Eleventh Conference on Applied Linguistics (CONAPLIN 2018)* (pp. 7-11). Atlantis Press. <https://doi.org/10.2991/conaplin-18.2019.216>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sneffjella, B., & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition*, 156, 135-146. <https://doi.org/10.1016/j.cognition.2016.07.010>
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press. <https://doi.org/10.5281/zenodo.3735822>
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell Publishers.
- Taylor, J. R. (2003). *Linguistic categorization* (3rd ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780199266647.001.0001>
- Zhang, W. (2009). Semantic prosody and ESL/EFL vocabulary pedagogy. *TESL Canada Journal*, 26(2), 1-12. <https://doi.org/10.18806/tesl.v26i2.411>