https://doi.org/10.26565/2073-4379-2025-47-03
UDC 33:81’373.612.2(336)

INTERDISCIPLINARY INTEGRATION OF ECONOMICS AND LINGUISTICS: CORPUS-BASED VOCABULARY MODELLING FOR FINANCIAL ENGLISH

Oleh Zatserkovnyi
Senior Teacher, the Department of Modern European Languages,
State University of Trade and Economics
(02156, Kyiv, 19 Kyoto Str.);
e-mail: o.zatserkovnyi@knute.edu.ua;
ORCID: https://orcid.org/0000-0003-3236-4418

This study addresses the growing need for linguistically grounded and domain-sensitive vocabulary modelling in Financial English by integrating methods from corpus linguistics with insights from financial discourse analysis. While financial communication is central to global economic systems, the vocabulary that structures this domain remains under-examined, particularly in terms of how lexical items function across different genres such as journalism, corporate reporting, and regulatory texts. This gap is especially evident in English for Specific Purposes (ESP) instruction, where existing materials often lack empirical grounding and fail to reflect the complexity of real-world financial language use. The purpose of the study is to develop a corpus-based vocabulary model that captures the semantic, collocational, and discourse functions of Financial English vocabulary. A 878,000-word Financial English Corpus (FEC) was constructed, incorporating 220 texts from financial journalism, corporate financial reports, and institutional regulatory documents. The methodological framework includes frequency analysis, mutual information (MI)-based collocational profiling, semantic tagging using the UCREL USAS system, and qualitative concordance analysis. The results demonstrate that core financial terms such as inflation, capital, risk, and yield exhibit genre-specific frequencies, collocational partners, and discourse functions. For instance, inflation is highly evaluative in regulatory discourse, while capital is linked to strategy in corporate texts. Semantic tagging revealed a significant presence of evaluative and affective language, challenging assumptions of objectivity in financial writing. The study concludes that a corpus-based model offers a replicable and pedagogically valuable approach to specialised vocabulary. The output has practical implications for ESP instruction, terminology management, and financial NLP, contributing to a more nuanced understanding of how financial meaning is constructed in institutional contexts.

Keywords: *corpus linguistics, discourse analysis, ESP, financial English, vocabulary modelling.*

Statement of the problem. Financial communication today operates at the intersection of linguistic precision and economic consequence, yet the vocabulary underpinning this domain remains systematically understudied. Despite the critical role of specialised terminology in shaping market cognition and professional discourse, Financial English lacks the corpus-driven models now standard in comparable fields such as legal or medical English [4; 6]. This gap reflects a deeper interdisciplinary divide: while corpus linguistics has evolved into a powerful tool for analysing authentic language use [2], its application to financial discourse remains limited. This research gap manifests bidirectionally. Economic scholarship typically treats language as a neutral transmission medium, overlooking how rhetorical structures, lexical framing, and metaphorical constructs actively influence financial decision-making and behaviour [15; 16]. Conversely, linguistic analyses of financial vocabulary often remain detached from the socio-economic functions these terms serve, limiting their relevance to practitioners and educators. Even within English for Specific Purposes (ESP) research – which emphasises discipline-sensitive pedagogy – consistent, empirically grounded models for financial vocabulary are notably absent [9].

Existing resources, such as financial glossaries and dictionaries, are typically compiled prescriptively rather than derived from large-scale authentic corpora. Recent corpus-driven research demonstrates that frequency analysis combined with collocation and concordance studies reveals usage patterns invisible to intuition-based methods [5], making them essential for capturing the contextual variability of financial terminology. Financial texts present particular analytical challenges: they deploy extensive hedging, evaluative lexis, and domain-specific metaphors that require both quantitative frequency analysis and qualitative discourse interpretation [8]. Recent developments in corpus-assisted approaches to specialised discourse have demonstrated the value of integrating computational methods with discourse analysis [17]. However, despite growing interest in financial discourse analysis [13], comprehensive vocabulary models that bridge linguistic description and economic domain knowledge remain scarce. Furthermore, the rise of digital finance and fintech has introduced new terminological challenges that existing frameworks have yet to address systematically [11].

Analysis of recent research and publications. To frame the present study, this literature review synthesises key developments in Corpus Linguistics applied to specialised registers, the discourse analysis of financial texts, and recent advances in vocabulary modelling techniques. These interconnected areas illuminate both the potential and the current gaps in corpus-based approaches to financial language.

Corpus linguistics has become a foundational methodology for analysing authentic language use in professional and institutional contexts. Baker [2] emphasises the role of corpora in identifying discourse patterns that elude introspection, particularly in domains such as politics, healthcare, and education. Expanding this approach, Gablasova, Brezina, and McEnery [5] demonstrate how collocational analysis, frequency profiling, and concordance searches provide empirical grounding for vocabulary instruction in English for Specific Purposes (ESP). Similarly, Goźdź-Roszkowski [6] shows how phraseological patterns reveal institutional preferences and discursive norms in legal and bureaucratic genres, underscoring the relevance of corpus-driven studies for domain-specific language. Kenny, Işık-Taş, and Jian [9] synthesise current innovations in ESP pedagogy, highlighting how learner-oriented corpora inform instructional design. Yet, financial English – despite being critical to global communication – remains underrepresented in this literature, especially in contrast to the well-established corpora and taxonomies available for medical or legal English. With regard to the financial discourse analysis, recent scholarship has increasingly turned its attention to the linguistic mediation of financial systems. Mooney [13] explores how financial texts construct economic ideologies, noting that discourse both reflects and shapes market behaviour. Jaworska [8] discusses the persuasive and identity-building functions of corporate language, highlighting how financial narratives manage risk, credibility, and authority. Tanweer and Avsar [15] focus on the ideographic use of economic terms, revealing how lexical choices can obscure complex financial realities. Metaphor remains a central concern in this strand of research. Telibaşa [16] documents the prevalence of metaphorical framing in economic texts, particularly in crisis contexts, while Lu (2023) investigates how fintech discourse incorporates hybrid metaphors to articulate technological innovation. Despite these insights, many of these studies adopt discourse-analytic or rhetorical approaches without integrating empirical corpus methods at scale.

Vocabulary modelling has benefited substantially from corpus methodologies that go beyond word lists to include semantic fields, collocational networks, and frequency bands. Gablasova et al. [5] advocate for multi-dimensional analyses that combine statistical and contextual data. Wu and Yue [17] note the growing relevance of corpus-assisted discourse studies (CADS), which merge quantitative tools with interpretive frameworks. Recent computational advances have opened new possibilities for domain-specific vocabulary modelling. Choe et al. [4] evaluate how corpus diversity affects the training of financial language models, indicating that richer corpora enhance semantic nuance in downstream tasks like sentiment analysis and named entity recognition. However, most modelling efforts in this area prioritize NLP performance rather than linguistic theory, leaving a methodological gap in interpretable, corpus-grounded vocabulary models that serve both scholarly and pedagogical ends. A consistent theme across this literature is the segregation of methodologies and disciplinary assumptions. Financial linguistics often relies on discourse analysis with limited corpus grounding, while computational modelling of financial texts rarely engages with linguistic theory. Furthermore, ESP research lacks comprehensive, replicable vocabulary frameworks for finance that can support teaching, translation, or automated processing.

The relevance of the paper lies in its contribution to current global demands for clearer, more effective financial communication across educational, professional, and technological domains. As financial information becomes increasingly accessible to non-experts through digital platforms, the ability to model financial vocabulary accurately is vital for enhancing financial literacy, designing more targeted English for Specific Purposes (ESP) curricula, and improving AI-driven financial text analysis. Moreover, the rapid evolution of financial technologies and the global spread of fintech discourse have introduced new terminologies and linguistic patterns that traditional glossaries or pedagogical tools struggle to capture. By offering a corpus-based, domain-aware vocabulary model, this study responds to the need for more adaptive, interpretable, and multilingual approaches to financial language in a globalised, digitised economy.

The ***purpose*** of this paper is to develop a corpus-based vocabulary model of Financial English by integrating methods from corpus linguistics with insights from financial discourse and underlying economic conceptual frameworks. The study seeks to create a

linguistically interpretable and domain-relevant model that not only enhances theoretical understanding of specialised financial language but also supports practical applications in education, translation, and natural language processing.

Presentation of the main material of the study. This study employed a mixed-methods corpus linguistics framework that integrates quantitative lexical analysis with qualitative discourse interpretation. Grounded in the principles of corpus-assisted discourse studies (CADS), the methodology supports the development of a vocabulary model that reflects both statistical regularities and functional meanings in Financial English [14; 17]. By aligning linguistic analysis with economic domain knowledge, the study aimed to produce a vocabulary model that is empirically grounded, contextually nuanced, and practically applicable. The analysis drew on a custom-compiled Financial English Corpus (FEC), totalling approximately eight hundred fifty thousand words. The corpus comprises three major genres of financial communication: journalistic texts – articles from The Financial Times, Bloomberg, and Reuters – corporate financial documents (such as annual reports and earnings calls), and regulatory or institutional publications (notably central bank statements and documents from international financial institutions like the IMF and World Bank). This genre-based approach follows recommendations from corpus design literature emphasising textual representativeness and diversity [2; 9]. Text processing was conducted using Sketch Engine [10] and AntConc [1], well-established tools in corpus linguistics research. The corpus was tokenised, lemmatised, and part-of-speech tagged using TreeTagger [12], and standard stopwords and non-lexical items were excluded. This preprocessing approach aligns with best practices in specialised corpus construction [3]. High-frequency terms were manually inspected to reduce lexical noise and improve data quality. Keyword extraction, collocation profiling, and frequency distribution analysis were carried out to identify salient vocabulary patterns [5]. Semantic tagging was selectively applied using the UCREL Semantic Analysis System (USAS), allowing lexical items to be grouped into key financial categories (risk, evaluation, modality). This step supports domain-sensitive vocabulary modelling, particularly in ESP and discourse analysis contexts [7; 13]. The vocabulary modelling proceeded in three stages. First, frequency thresholding identified lexical items that appeared with statistically significant

frequency (per million words) across at least two of the three genres. Second, collocational behaviour was analysed using Mutual Information (MI) and log-likelihood ratios, which are widely used to detect significant lexical associations [3; 5]. Third, a qualitative discourse analysis of concordance lines was conducted to explore evaluative, metaphorical, and argumentative uses of key vocabulary – a strategy common in CADS and financial discourse studies [8; 11; 16]. To integrate economic expertise, key terms were mapped against standard classifications from sources such as the OECD and international accounting frameworks. Selected findings were reviewed with finance professionals to ensure conceptual validity and relevance to domain-specific practices. The final output included a frequency-ranked vocabulary list, annotated with common collocates, semantic tags, and discourse functions. Validation involved triangulation across subcorpora, independent double coding of discourse functions, and pilot testing in an ESP course on Financial English, consistent with current recommendations for applied corpus studies in educational contexts [7; 9]. This methodology ensured the resulting model was linguistically interpretable, statistically robust, and pedagogically useful for applications in financial literacy, translation, and natural language processing.

Corpus Overview

The Financial English Corpus (FEC) compiled for this study comprises approximately 878,000 words drawn from a carefully selected set of 220 texts across three core financial genres: financial journalism (350,000 words), corporate financial reporting (330,000 words), and regulatory/institutional discourse (198,000 words) (see Table 1). Texts were sourced from prominent financial communication channels between 2016 and 2024, including *The Financial Times*, *Bloomberg*, annual reports from global financial institutions (*BlackRock, JP Morgan*), and central bank statements (*ECB, Federal Reserve*). While smaller than general-purpose corpora, the FEC meets the methodological threshold for robust analysis of medium-to-high frequency vocabulary, collocational patterns, and genre-based lexical variation, particularly within the scope of English for Specific Purposes (ESP) and corpus-assisted discourse studies (CADS). The balanced genre distribution and temporal coverage enable reliable cross-sectional analysis while maintaining sufficient depth for domain-specific modelling.

Table 1. Composition of the Financial English Corpus (FEC) by Genre, Text Count, Token Count, and Date Range.

Genre	*No. of Texts*	*Tokens*	*Date Range*
Financial Journalism	90	350,000	2016–2024
Corporate Reports	65	330,000	2015–2023
Regulatory Discourse	65	198,000	2016–2024
Total	**220**	**878,000**	

The corpus findings yield several critical insights into the nature of Financial English and the utility of corpus-driven methods for vocabulary modelling. First, the analysis confirms that financial vocabulary is semantically dense and functionally multidimensional. High-frequency terms such as market, inflation, volatility, and capital operate not only as technical referents but also as vehicles of evaluation, hedging, and strategic framing, especially in journalistic and regulatory discourse. This underscores the importance of modelling financial vocabulary not just in terms of statistical frequency, but also in relation to discourse function and semantic role. Second, the study demonstrates that a mid-scale, domain-targeted corpus – such as the 878,000-word FEC – is sufficient for extracting statistically meaningful lexical and collocational data when paired with genre-sensitive design. The corpus enabled the identification of core terms, genre-specific collocations, and consistent discourse patterns across financial communication channels. This validates the methodological premise that focused corpus design, rather than corpus size alone, is key to high-resolution vocabulary modelling in ESP contexts. Third, the corpus output provides a functional vocabulary model that can be directly applied to several use cases:

- In ESP curriculum design, it offers a data-driven basis for teaching core financial terminology and usage patterns.
- For translation and terminology management, it supports the development of contextualised glossaries sensitive to semantic shifts across genres.
- In financial NLP, the annotated vocabulary – particularly collocational frames and discourse roles – can enhance the performance and interpretability of language models used in sentiment analysis, risk assessment, or automated financial reporting.

Despite its analytical yield, the corpus also imposes certain limitations. Its size restricts exploration of low-frequency neologisms,

such as emerging ESG or crypto-finance terminology. Moreover, diachronic trends are not fully analysable without longitudinal subcorpora. To address these gaps, future research should consider expanding the FEC with digital financial discourse (fintech blogs, investor forums, social media) to capture linguistic innovation and the evolving rhetoric of finance in real time.

Frequency-Based Core Vocabulary

To identify the lexical backbone of Financial English, a frequency threshold of 50 occurrences per million words was applied across the corpus. This yielded a list of 112 core financial terms that appeared in at least two out of the three genres, suggesting a shared lexicon underlying financial discourse. These terms include foundational economic concepts such as market, risk, capital, and interest, alongside evaluative and policy-laden vocabulary such as inflation, volatility, and growth. The table below presents a representative subset of these terms and their normalised frequency (per million words) in each genre. The variation across genres highlights how the prominence of certain terms aligns with institutional function and communicative goals.

Table 2. Normalised Frequency (Per Million Words) of Selected Core Financial Terms Across Genres in the FEC.

Term	*Journalism (per mil)*	*Corporate Reports*	*Regulatory Discourse*
market	385	310	275
inflation	112	88	346
capital	152	277	134
volatility	97	65	102
yield	63	54	145

The results in Table 2 demonstrate that even within a shared disciplinary domain, lexical distribution is genre-sensitive, shaped by the communicative priorities and institutional roles of different text types. The term market appears frequently across all three genres, reinforcing its status as a core concept in financial discourse. However, other terms – such as capital, inflation, and yield – exhibit significant variation across contexts.

For instance, capital occurs most frequently in corporate financial reports, where it is central to discussions of equity structure, funding strategies, and investment flows. In contrast, inflation is most prevalent in regulatory discourse, reflecting the macroeconomic responsibilities of central banks and international institutions. Meanwhile, yield

appears notably in regulatory and journalistic texts, consistent with its relevance in policy announcements and financial commentary on bond markets. These differences are not merely quantitative – they signal the functional specialisation of vocabulary within financial sub-genres. In regulatory contexts, terms like inflation and yield are used to frame monetary policy narratives, while in journalism, the same terms appear in market interpretation and forecasting. In corporate reports, vocabulary like capital, return, and liquidity tends to be used in the context of strategic positioning and performance disclosure.

The frequency patterns (in Table 2) directly support the purpose of the study: to build a linguistically interpretable and domain-relevant vocabulary model of Financial English. The data shows that high-frequency financial terms cannot be fully understood outside of their genre context. Their meanings, collocational behaviour, and discourse roles shift depending on who is using them, and for what communicative purpose.

This has several important implications:

- For ESP education, these findings show that teaching "financial vocabulary" in isolation is insufficient. Vocabulary instruction should be genre-specific and include usage contexts.

- For computational applications, such as NLP and financial text mining, genre-sensitive frequency profiling can improve the accuracy of tasks like named entity recognition, topic modelling, and sentiment analysis.

- For translation and terminology development, the variation in frequency and function underscores the need to align financial terms with their typical collocational environments and genre-specific meanings.

In short, these findings validate the genre-aware and corpus-driven approach taken in this study. They demonstrate that a vocabulary model of Financial English must go beyond a simple frequency list to reflect how financial terms are actually distributed, framed, and used across professional contexts. This corpus-based evidence lays the groundwork for building a more dynamic, functional, and pedagogically useful vocabulary model.

Collocational Patterns in Financial English

A central goal of this study is to model Financial English in a way that captures not just individual term frequency, but also how terms operate within fixed or semi-fixed collocational frames. Collocational analysis provides crucial insight into the lexico-grammatical behaviour of high-frequency terms and reveals how domain knowledge is

linguistically structured through recurring lexical pairings. Using Mutual Information (MI) scores to identify statistically significant collocates, the study identified key lexical partnerships that form the building blocks of financial discourse. Table 3 presents representative results for three core terms – *interest*, *risk*, and *yield* – including their most salient collocates and MI score ranges. These values were extracted from the full corpus and cross-checked by genre to observe contextual variation.

Table 3. High-Strength Collocational Frames for Core Financial Terms in the FEC
(MI Score ≥ 6.0).

Node Word	*Top Collocates*	*MI Score Range*
interest	rate, hike, cut, policy	6.3–9.8
risk	management, appetite, exposure	7.1–8.5
yield	curve, bond, treasury	6.9–10.2

The collocational behaviour of these terms reveals genre-specific and function-driven lexical patterns. The node word *interest* consistently forms bundles such as interest rate hike, interest rate decision, and monetary policy cut, which are particularly frequent in journalistic and regulatory texts. These frames function as predictive or evaluative structures, often used to interpret macroeconomic signals or anticipate central bank actions. The term *risk* reveals a different collocational ecosystem, with frequent pairings like risk management and risk appetite, which dominate corporate reports. These collocations reflect institutional discourse around compliance, internal control, and investor strategy, highlighting the corporate sector's focus on risk mitigation and disclosure. *Yield* co-occurs with terms like curve, bond, and treasury, primarily in regulatory and investment-related texts. These collocates not only describe asset performance but also help construct market sentiment narratives – for example, inverted yield curve is a known recession signal, and its appearance in concordance lines often marks speculative or cautionary discourse. What these patterns reveal is that collocational strength is not merely a stylistic feature, but a semantic and pragmatic signal of how financial concepts are contextualised and deployed. These lexical partnerships encode domain expertise and are essential for genre competence in financial communication.

These findings directly serve the study's central purpose: to create a domain-sensitive vocabulary model of Financial English. The analysis

shows that financial expertise is communicated through collocational precision, not just individual word choice. Understanding how key terms like interest, risk, or yield behave lexically enables the construction of a vocabulary model that goes beyond surface frequency to capture the institutional logic and discourse function embedded in financial texts.

This level of granularity is particularly valuable for:

- ESP instruction: Learners benefit from exposure to authentic phraseology, not isolated vocabulary items.
- Terminology extraction and translation: Recognising frequent collocates improves contextual accuracy and helps avoid semantic drift.
- Financial NLP and information retrieval: Systems trained on surface word frequency miss these semantic pairings, whereas collocational models improve both precision and interpretability.

Thus, the collocational analysis is not a supplementary finding – it is a foundational component of the vocabulary model proposed by this study, confirming that financial terminology operates within highly regularised, genre-sensitive lexical frameworks that must be captured for any meaningful applied or theoretical use.

Discourse Functions and Semantic Roles

Beyond frequency and collocation, one of the key aims of this study is to uncover how financial vocabulary operates discursively – that is, how terms contribute not only to content delivery but to the framing, evaluation, and interpretation of financial phenomena. To this end, concordance analysis and semantic tagging were used to examine the discourse functions and pragmatic roles of high-frequency financial terms. The concordance analysis revealed that financial terms frequently occur in modalised or evaluative constructions, particularly in journalistic and regulatory genres. Phrases such as "*may signal risk*", "*expected to rebound*", and "*projected growth*" are not simply descriptive – they are forms of linguistic forecasting and hedging that frame uncertainty, imply institutional authority, or anticipate market behaviour. Semantic tagging using the UCREL Semantic Analysis System (USAS) further substantiated these patterns. The dominant semantic categories across the corpus were:

- F1 (money and pay) – capturing core financial entities and transactions.
- X4 (evaluation) – indicating judgement, appraisal, and quality assessment.
- A2 (affect/attitude) – reflecting emotional tone, confidence, or concern.

Table 4. Dominant Semantic Fields in Financial English by Token Frequency.

Semantic Category	*% of Tagged Tokens*	*Typical Examples*
F1 (Finance)	28.11	inflation, yield
X4 (Evaluation)	19.53	strong, weak
A2 (Affect)	11.71	concern, confidence

As can be seen in Table 4, the dominance of F1 tokens confirms the expected centrality of monetary terminology in financial texts. However, what is more revealing is the high frequency of X4 (evaluation) and A2 (affect) tags. These categories indicate that financial texts are not merely objective vehicles of numerical reporting, but are linguistically structured to manage perception, credibility, and expectation. For example, the use of terms like strong growth, weak demand, or market concern demonstrates how language in financial contexts is laden with strategic evaluation. This is particularly pronounced in journalism and regulatory texts, where tone, caution, and optimism are not neutral but instrumental in guiding investor sentiment and public trust. Moreover, the presence of affective language (confidence, anxiety, uncertainty) in corporate disclosures and institutional reports suggests that emotional calibration plays a subtle but crucial role in how organisations manage risk narratives and public accountability. These rhetorical choices are part of what constructs institutional voice and authority in financial communication.

These findings are essential to the study's overarching goal: to develop a corpus-based vocabulary model of Financial English that is not only statistically representative but functionally and pragmatically grounded. Understanding how financial terms are used to evaluate, hedge, or project meaning allows the vocabulary model to move beyond wordlists and into real-world communicative function.

This matters for several reasons:

- For ESP instruction, it highlights the need to teach not just terminology, but the discursive strategies that professionals use to signal uncertainty, align with institutional tone, or frame financial outcomes.

- For translators and lexicographers, it underscores the necessity of capturing functional equivalence, not just semantic parity.

- For NLP developers, recognising the affective and evaluative layers of financial discourse is critical for improving sentiment analysis, text classification, and risk language detection in financial automation tools.

In short, these discourse-level insights reinforce that financial vocabulary is not used in isolation. It is embedded in predictable

rhetorical patterns that vary by genre and communicative purpose – patterns that any effective vocabulary model must account for.

Vocabulary Model Output

The culminating product of this study is a structured vocabulary model of Financial English that integrates corpus frequency analysis, collocational profiling, and discourse-level interpretation. Unlike traditional wordlists or glossaries, this model captures how financial terminology functions contextually across genres, genres, and rhetorical purposes – precisely aligning with the study's core aim: to develop a linguistically and domain-informed vocabulary framework that supports both theoretical inquiry and applied pedagogy.

The model is built on a three-part structure (see Table 5):

1) *Core Terms:* Lexical items with normalised high frequency and broad genre coverage (market, capital, inflation).

2) *Collocational Frames*: Genre-sensitive multiword expressions and lexical bundles (interest rate hike, risk appetite, market volatility).

3) *Discourse Functions:* Annotated semantic-pragmatic roles derived from concordance data (evaluation, hedging, forecasting).

Table 5. Illustrative Vocabulary Model Output with Semantic and Discourse Annotation.

Term	*Collocational Frame*	*Genre Dominance*	*Semantic Tag*	*Discourse Role*
inflation	inflation pressure, rate	Regulatory	F1	Evaluation/Warning
capital	raise capital, capital flow	Corporate	F1	Strategic Action
volatility	market volatility, spike	Journalism	X4	Risk Framing

This vocabulary model does more than classify words by frequency – it captures the functional architecture of financial communication. For instance, inflation is not just a high-frequency term; it functions predominantly in regulatory texts as an indicator of macroeconomic stability, often paired with hedging or cautionary discourse. Similarly, capital in corporate reports appears within collocations related to financing and growth strategy, reflecting institutional goals and stakeholder positioning. Volatility is leveraged in journalistic contexts to construct narratives of market instability and urgency. The semantic tagging and discourse role annotation provide a robust layer of interpretation, allowing the model to distinguish

between literal, metaphorical, and evaluative uses of vocabulary. This nuance is essential for learners, translators, and NLP systems seeking to move beyond word-level analysis into meaningful contextual understanding. From a pedagogical standpoint, this model is transformative for English for Specific Purposes (ESP) instruction in the financial domain. Unlike traditional ESP wordlists, which often isolate terminology from its usage environment, this model presents learners with:

- Authentic lexical bundles as used in real-world financial genres.
- Insights into how vocabulary varies across institutional discourse types.
- Exposure to pragmatic functions like hedging, forecasting, and strategic framing.

These features enable instructors to move from vocabulary acquisition to genre competence, equipping learners with the tools to interpret and produce texts in professional financial settings. This also supports materials development for ESP textbooks, glossaries, and task-based learning activities grounded in corpus evidence.

This vocabulary model contributes to applied corpus linguistics, financial discourse analysis, and ESP research in several ways:

- It demonstrates a replicable methodology for modelling specialised vocabulary through corpus tools and discourse annotation.
- It reveals how financial concepts are linguistically operationalized in different institutional contexts, filling a gap in prior research that often treated financial vocabulary as static or decontextualised.
- It bridges the disciplinary divide between linguistics and economics, offering a data-driven foundation for interdisciplinary research in communication, education, and financial technologies (explainable NLP in fintech).

In sum, this model not only serves as a resource for teaching and translation, but also offers a scalable research tool for future work in specialized language modelling and domain-specific semantics.

Conclusions. This study set out to develop a corpus-based vocabulary model of Financial English that reflects both linguistic structure and economic domain knowledge. Through the construction and analysis of a genre-balanced 878,000-word Financial English Corpus (FEC), the research achieved its objective by combining frequency profiling, collocational analysis, and semantic-discourse annotation to map the lexical architecture of financial discourse. The

findings show that Financial English is not a monolithic register but a network of genre-sensitive lexico-semantic patterns, shaped by institutional function and communicative intent. Core financial terms – such as market, inflation, capital, and risk – exhibit distinct collocational frames and discourse roles depending on whether they appear in journalistic, corporate, or regulatory contexts. Semantic tagging further revealed a consistent presence of evaluative and affective language, confirming that financial texts are strategically framed, not merely descriptive.

The resulting vocabulary model captures these complexities by integrating core terms, collocational frames, and discourse functions, offering a dynamic and interpretable structure for use in pedagogy, translation, and computational applications. For ESP instruction, the model provides a foundation for genre-aware materials that reflect authentic financial language in use. For linguists and language technologists, it demonstrates how corpus methods can bridge the gap between formal lexical data and pragmatic, real-world meaning. Ultimately, the study contributes to both the theoretical understanding of financial discourse and the practical modelling of specialised vocabulary in applied linguistics. It also lays groundwork for further interdisciplinary exploration between corpus linguistics and economic communication studies.

Perspectives for further research. Future studies should expand the corpus with more diverse and digital financial texts to capture evolving vocabulary and diachronic trends. The vocabulary model should be tested in ESP classrooms to assess its pedagogical value and adapted for integration into NLP systems for improved domain-specific language processing. Cross-linguistic comparisons could also enhance its relevance for multilingual financial communication and translation.

REFERENCES

1. Anthony, L. (2023). *AntConc* (Version 4.2.0) [Computer software]. Waseda University. Available at: ttps://www.laurenceanthony.net/software/ antconc/ [Accessed 13 July 2025].
2. Baker, P. (2023). *Using Corpora in Discourse Analysis* (2nd ed.). Bloomsbury Academic. Available at: https://www.bloomsbury.com/uk/ using-corpora-in-discourse-analysis-9781350083752/ [Accessed 13 July 2025].
3. Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. DOI: https://doi.org/10.1017/9781316410899

4. Choe, J., Noh, K., Kim, N., Ahn, S., Jung, W. (2023). Exploring the Impact of Corpus Diversity on Financial Pretrained Language Models. *arXiv.Org*. abs/2310.13312. DOI: https://doi.org/10.48550/arxiv.2310.13312
5. Gablasova, D., Brezina, V., McEnery, A.M. (2017). Collocations in corpus-based language learning research: Identifying, comparing and interpreting the evidence. *Language Learning. 67 (Suppl. 1)*, pp. 155-179. DOI: https://doi.org/10.1111/lang.12225
6. Goźdź-Roszkowski, S. (2019). *Phraseology in Legal and Institutional Settings.* Routledge.
7. Hyland, K. (Ed.). (2016). *The Routledge Handbook of English for Academic Purposes* (1st ed.). Routledge. DOI: https://doi.org/10.4324/9781315657455
8. Jaworska, S. (2020). Corporate discourse. In: De Fina, A. and Georgakopoulou-Nunes, A. (Eds.). *Handbook of Discourse Studies.* Cambridge University Press, pp.1-28. DOI: https://doi.org/10.1017/9781108348195
9. Kenny, N., Işık-Taş, E.E., Jian, H. (2020). *English for Specific Purposes Instruction and Research: Current Practices, Challenges and Innovations.* Cham: Palgrave Macmillan. DOI: https://doi.org/10.1007/978-3-030-32914-3
10. Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography. 1(1),* pp. 7-36. DOI: https://doi.org/10.1007/s40607-014-0009-9
11. Lu, L. (2023). Chapter 19: Fintech: technology-enabled financial innovation for digital trade. *Research Handbook on Digital Trade*. Cheltenham, UK: Edward Elgar Publishing, pp. 306-328. DOI: https://doi.org/10.4337/9781800884953.00029
12. Màrquez, L., Rodríguez, H. (1998). Part-of-speech tagging using decision trees. In: Nédellec, C., Rouveirol, C. (Eds.). *Machine Learning: ECML-98. ECML 1998. Lecture Notes in Computer Science*. Springer, Vol. 1398. DOI: https://doi.org/10.1007/BFb0026668
13. Mooney, A. (2020). The Discourses of Money and the Economy. In A. De Fina, A. Georgakopoulou (Eds.). *The Cambridge Handbook of Discourse Studies.* Cambridge: Cambridge University Press, pp. 644-665.
14. Partington, A., Duguid, A., Taylor, C. (2013). Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS). In G. Brookes (Ed.). *International Journal of Corpus Linguistics. 19(2),* pp. 292-300. DOI: https://doi.org/10.1075/ijcl.19.2.07bro
15. Tanweer, A.D., Avsar, R.B. (2015). Ideographic use of economic terms. *On the Horizon. 23(3),* pp. 169-173. DOI: https://doi.org/10.1108/OTH-05-2015-0018
16. Telibașa, G. (2015). The pervasiveness of metaphor in the language of economics. *Studies and Scientific Researches. Economics Edition. 21*. DOI: https://doi.org/10.29358/sceco.v0i21.305

17. Wu, X., Yue, X. (2025). Corpus-assisted discourse studies. In M. Gillings, G. Mautner, P. Baker (Eds.). *Digital Scholarship in the Humanities. 40(1),* pp. 457-459. DOI: https://doi.org/10.1093/llc/fqae066

The article was received by the editors 14.09.2025.
The article was recommended for printing 17.11.2025.
The article published 30.12.2025.

In cites: Zatserkovnyi O. (2025). Interdisciplinary integration of economics and linguistics: Corpus-based vocabulary modelling for Financial English. *Teaching languages at higher educational establishments at the present stage. Intersubject relations. 47,* pp. 42-60. https://doi.org/10.26565/2073-4379-2025-47-03

МІЖДИСЦИПЛІНАРНА ІНТЕГРАЦІЯ ЕКОНОМІКИ ТА ЛІНГВІСТИКИ: КОРПУСНЕ МОДЕЛЮВАННЯ ЛЕКСИКИ ФІНАНСОВОЇ АНГЛІЙСЬКОЇ МОВИ

Олег Зацерковний
старший викладач кафедри сучасних європейських мов
Державного торговельно-економічного університету
(02156, Київ, вул. Кіото, 19);
e-mail: o.zatserkovnyi@knute.edu.ua;
ORCID: https://orcid.org/0000-0003-3236-4418

У цьому дослідженні розглядається зростаюча потреба в лінгвістично обґрунтованому та чутливому до предметної сфери моделюванні лексики фінансової англійської мови шляхом поєднання методів корпусної лінгвістики з підходами фінансового дискурс-аналізу. Незважаючи на те, що фінансова комунікація є ключовим складником глобальної економічної системи, лексика, що структурує цей дискурс, досі недостатньо досліджена – особливо з точки зору того, як лексичні одиниці функціонують у різних жанрах, таких як журналістика, корпоративна звітність та регуляторні тексти. Цей розрив особливо помітний у викладанні англійської мови за професійним спрямуванням (ESP), де наявні матеріали часто не мають емпіричного підґрунтя і не відображають складності реальної комунікації у сфері фінансів. Метою дослідження є розробка корпусної моделі лексики фінансової англійської, яка охоплює її семантичні, колокаційні та дискурсивні функції. Було створено корпус Financial English Corpus (FEC) обсягом 878 000 слів, до якого увійшли 220 текстів фінансової журналістики, корпоративної звітності та регуляторних документів. Методологічна основа включає частотний аналіз, профілювання колокацій на основі взаємної інформації (MI), семантичне тегування за допомогою системи UCREL USAS та якісного конкорданс-аналізу. Результати показують, що ключові фінансові

терміни, зокрема *inflation, capital, risk* та *yield*, мають жанрову специфіку щодо частотності, колокаційних партнерів та дискурсивних функцій. Наприклад, *inflation* має чітко оцінний характер у регуляторному дискурсі, тоді як *capital* пов'язується зі стратегічною комунікацією в корпоративному контексті. Семантичне тегування виявило значну присутність оцінної та афективної лексики, що кидає виклик традиційному уявленню про об'єктивність фінансового письма. Дослідження доходить висновку, що модель, побудована на корпусному аналізі, пропонує відтворюваний і дидактично цінний підхід до спеціалізованої лексики. Отриманий результат має практичне значення для викладання ESP, управління термінологією та розробки систем автоматичної обробки фінансових текстів (NLP), сприяючи глибшому розумінню того, як формується фінансове значення в інституційних контекстах.

***Ключові слова**: англійська для спеціальних цілей, дискурс-аналіз, корпусна лінгвістика, моделювання лексики, фінансова англійська.*

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Anthony L. *AntConc: Version 4.2.0* [Електронний ресурс]. Комп'ютерна програма. Waseda University, 2023. URL: https://www.laurenceanthony.net/software/antconc/ (дата звернення 13.07.2025).
2. Baker P. *Using Corpora in Discourse Analysis.* 2nd ed. London: Bloomsbury Academic, 2023. URL: https://www.bloomsbury.com/uk/ using-corpora-in-discourse-analysis-9781350083752/ (дата звернення 13.07.2025).
3. Brezina V. *Statistics in corpus linguistics: A practical guide.* Cambridge: Cambridge University Press, 2018. DOI: https://doi.org/10.1017/9781316410899
4. Choe J., Noh K., Kim N., Ahn S., Jung W. Exploring the Impact of Corpus Diversity on Financial Pretrained Language Models. *arXiv.Org.* 2023. DOI: https://doi.org/10.48550/arxiv.2310.13312
5. Gablasova D., Brezina V., McEnery A.M. Collocations in corpus-based language learning research: Identifying, comparing and interpreting the evidence. *Language Learning.* 2017. Vol. 67, Suppl. 1. P. 155-179. DOI: https://doi.org/10.1111/lang.12225
6. Goźdź-Roszkowski S. Phraseology in Legal and Institutional Settings. London: Routledge, 2019.
7. Hyland K. (Ed.). The Routledge Handbook of English for Academic Purposes. 1st ed. London: Routledge, 2016. DOI: https://doi.org/10.4324/9781315657455
8. Jaworska S. Corporate discourse. In: De Fina A., Georgakopoulou-Nunes A. (Eds.). *Handbook of Discourse Studies*. Cambridge: Cambridge University Press, 2020. P. 1-28. DOI: https://doi.org/10.1017/9781108348195

9. Kenny N., Işık-Taş E.E., Jian H. *English for Specific Purposes Instruction and Research: Current Practices, Challenges and Innovations*. Cham: Palgrave Macmillan, 2020. DOI: https://doi.org/10.1007/978-3-030-32914-3
10. Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovář V., Michelfeit J., Rychlý P., Suchomel V. The Sketch Engine: Ten years on. *Lexicography*. 2014. Vol. 1, No. 1. P. 7-36. DOI: https://doi.org/10.1007/s40607-014-0009-9
11. Lu L. Fintech: technology-enabled financial innovation for digital trade. In: *Research Handbook on Digital Trade*. Cheltenham, UK: Edward Elgar Publishing, 2023. P. 306-328. DOI: https://doi.org/10.4337/9781800884953.00029
12. Màrquez L., Rodríguez H. Part-of-speech tagging using decision trees. In: Nédellec C., Rouveirol C. (Eds.). *Machine Learning: ECML-98. Lecture Notes in Computer Science*. Vol. 1398. Berlin: Springer, 1998. DOI: https://doi.org/10.1007/BFb0026668
13. Mooney A. The Discourses of Money and the Economy. In: De Fina A., Georgakopoulou A. (Eds.). *The Cambridge Handbook of Discourse Studies*. Cambridge: Cambridge University Press, 2020. P. 644-665.
14. Partington A., Duguid A., Taylor C. Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS). In: Brookes G. (Ed.) *International Journal of Corpus Linguistics*. 2013. Vol. 19, No. 2. P. 292-300. DOI: https://doi.org/10.1075/ijcl.19.2.07bro
15. Tanweer A.D., Avsar R.B. Ideographic use of economic terms. *On the Horizon*. 2015. Vol. 23, No. 3. P. 169-173. DOI: https://doi.org/10.1108/OTH-05-2015-0018
16. Telibașa G. The pervasiveness of metaphor in the language of economics. *Studies and Scientific Researches. Economics Edition.* 2015. No. 21. DOI: https://doi.org/10.29358/sceco.v0i21.305
17. Wu X., Yue X. Corpus-assisted discourse studies. In: Gillings M., Mautner G., Baker P. (Eds.). *Digital Scholarship in the Humanities*. 2025. Vol. 40, No. 1. P. 457-459. DOI: https://doi.org/10.1093/llc/fqae066

Стаття надійшла до редакції 14.09.2025.
Статтю рекомендовано до друку 17.11.2025.
Стаття опублікована 30.12.2025.

Як цитувати: Зацерковний О. Міждисциплінарна інтеграція економіки та лінгвістики: корпусне моделювання лексики фінансової англійської мови. *Викладання мов у вищих навчальних закладах освіти на сучасному етапі. Міжпредметні зв'язки.* 2025. Вип. 47. С. 42-60. https://doi.org/10.26565/2073-4379-2025-47-03