

A corpus-based multi-paradigmatic analysis of multidisciplinary terms for ESP pedagogy: Cross-linguistic insights for Ukrainian tertiary education

Svitlana Terekhova ^{a,1*}, Olesia Markevych ^{b,2}, Olena Semenyshyn ^{3,c}, Liliia Ovcharenko ^{d,4},
Oksana Klak ^{e,5}

^a Kyiv National Linguistic University, Velyka Vasylykivska St, 73, Kyiv, Ukraine

^b State University of Trade and Economics, 19 Kyoto St, Kyiv, 03156, Ukraine

^c State Tax University, 31 Universytetska Street, Irpin, 08205, Kyiv Region, Ukraine

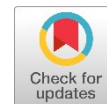
^d Cooperative Institute of Business and Law, Yulii Zdanovska St, 18, Kyiv, Ukraine

^e Drohobych Ivan Franko State Pedagogical University, Ivana Franka St, 24, Drohobych, Lviv Oblast, Ukraine

^f Lviv State University of Internal Affairs, 26 Horodotska Street, Lviv, Lviv Oblast, Ukraine

¹svitlanasute@gmail.com*, ²olesiabodunovastu@gmail.com, ³olenasemenyshynicbl@gmail.com, ⁴ovcharenko.liliia@gmail.com, ⁵pelyhksju@ukr.net

*Corresponding author



ARTICLE INFO

Article history

Received 23 February 2026

Revised 2 September 2026

Accepted 3 September 2026

Available Online 29 September 2026

Keywords

Multi-paradigmatic

Multirespectivity

Terminological variability

Term unification

Language consistency

ABSTRACT

The study of polyparadigmatic analysis of terms is timely due to the growth of interdisciplinary communication, which requires a deep understanding of how terms adapt across different fields of knowledge. The article aims to conduct a polyparadigmatic analysis of multidisciplinary terms in Ukrainian and English in order to establish their semantic, structural, and functional characteristics. The research utilized UberText 2.0, ukTenTen, the Corpus of Contemporary American English, and the Open Parallel Corpus. Quantitative, corpus-based, contextual, and definitional analyses were applied. The results were analyzed using descriptive statistics, the χ^2 test, cluster analysis, ANOVA, and Tukey's test. The findings showed that terms exhibit varying degrees of semantic flexibility and cross-disciplinary adaptation. The term "Field" showed the highest average correlation (0.82), indicating its universality, while "Agent" and "Vector" showed low correlation (0.26), indicating a narrow specialization. The terms "Code" (0.91) and "Structure" (0.75) had a high level of cross-linguistic equivalence, confirming their stability in translation. Cluster analysis identified three main groups: technical terms (average frequency 320), universal terms (210), and bio-technical terms (180). The study revealed significant semantic flexibility in multidisciplinary terms, confirming their disciplinary specificity and cross-linguistic consistency. Pedagogically, these findings provide ESP educators and curriculum designers with empirical evidence to prioritize vocabulary instruction based on cross-disciplinary frequency and semantic stability. The identified clusters offer a roadmap for designing corpus-informed lexical modules that reduce conceptual transfer errors and foster academic literacy. Future classroom-based research is needed to test the effectiveness of these terminological clusters in Ukrainian tertiary ESP settings.



© The Authors 2026. Published by Universitas Ahmad Dahlan.
This is an open access article under the [CC-BY-SA](#) license.



How to Cite: Terekhova, S., Markevych, O., Semenyshyn, O., Ovcharenko, L. & Klak, O. (2026). A corpus-based multi-paradigmatic analysis of multidisciplinary terms for ESP pedagogy: Cross-linguistic insights for Ukrainian tertiary education. *English Language Teaching Education Journal*, 9(2), 352-367. <https://doi.org/10.12928/eltej.v9i2.15890>

1. Introduction

Modern scientific discourse is characterized by intensive integration processes that significantly influence the development of terminology across various fields of knowledge. The rise of multidisciplinary facilitates the migration of terms across domains, resulting in both semantic variation and structural transformation of these units. In this context, the analysis of multidisciplinary terms becomes a crucial tool for understanding not only their essence but also the mechanisms of interdomain knowledge exchange (Hong, 2023). This analysis gains particular relevance with the integration of Ukrainian education into global academic markets (Ishchenko et al., 2022)

Multidisciplinary terms, by forming cross-domain connections, play a key role in generating new knowledge (Yukhymenko et al., 2024; Garcia et al., 2025). They function not only as linguistic units but also as complex phenomena integrating semantic, structural, and functional aspects. Their meanings vary depending on the context, necessitating a polyparadigmatic analysis (Kolmos et al., 2025). This approach enables the identification of regularities in these changes and outlines their contribution to the evolution of specific fields of knowledge.

Despite significant attention to terminology in modern linguistics, the semantic and structural transformations of multidisciplinary terms across different languages remain underexplored. The peculiarities of their functioning in Ukrainian and English, as well as the influence of these languages on global knowledge exchange, require more in-depth investigation (Djumambetova, 2023; Sterner, 2022). This study focuses on analyzing multidisciplinary terms in terms of their semantic variability, structural characteristics, and functional significance in scientific discourse. Particular attention is paid to the polyparadigmatic analysis of Ukrainian and English terminology to identify cross-linguistic features and patterns.

The multi-paradigmatic approach provides a comprehensive study of linguistic phenomena through the integration of methods from various scientific paradigms, including structural linguistics, cognitive linguistics, and sociolinguistics. It is especially effective for analyzing multidisciplinary terms, as it allows for simultaneously considering their formal structure, semantic transformations, societal and cultural aspects affecting their use, and more.

Figure 1 presents a diagram of the multi-paradigmatic approach to the analysis of multidisciplinary terms (Terekhova, 2025a). It illustrates the interaction of three key paradigms — linguistic, cultural, and methodological — used to ensure a holistic analysis.

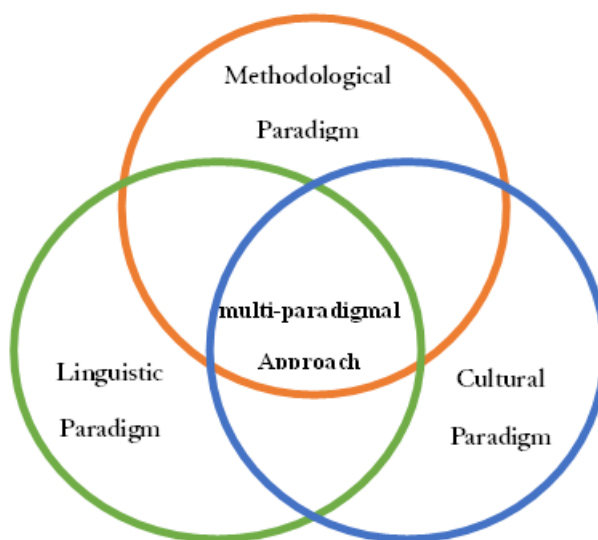


Fig 1. Visualization of the Multi-paradigmatic Approach

Source: Created by the authors.

The diagram illustrates the interaction of the three main paradigms within the multi-paradigmatic approach. The linguistic paradigm (blue circle) encompasses language structures, semantics, and syntax. The cultural paradigm (green circle) reflects the influence of cultural contexts on the meaning and usage of terms. The methodological paradigm (orange circle) emphasizes the disciplinary research methods of

terminology. At the intersection of these circles lies the field representing the multiparadigmatic as an integration of the mentioned paradigms to enable a comprehensive analysis (Terekhova, 2025b).

The novelty of the study lies in the comprehensive application of the multi-paradigmatic approach to identify patterns of semantic transformation of multidisciplinary terms in contemporary Ukrainian and English scientific discourse. For the first time, terms have been clustered according to their semantic flexibility and disciplinary specificity, allowing the distinction between their universal and context-dependent features. Crucially, this study bridges a critical gap between terminological research and English Language Teaching (ELT) by translating corpus-driven findings into actionable pedagogical strategies. While previous studies have described semantic shifts in terminology, they have not addressed how these shifts affect Ukrainian tertiary-level learners' academic literacy or ESP vocabulary acquisition. Specifically, this research addresses the challenges of vocabulary development, corpus-informed language instruction, and the prevention of conceptual transfer errors—issues that are central to ESP pedagogy but remain underexplored in Ukrainian higher education. By offering empirically grounded criteria for selecting and sequencing multidisciplinary terms in ESP syllabi, this study contributes not only to terminology science but also to the applied field of English for Specific Purposes.

The goal is to conduct a multi-paradigmatic analysis of multidisciplinary terms in Ukrainian and English to identify their semantic, structural, and functional characteristics. Objectives/questions:

1. To identify multidisciplinary terms across selected scientific domains in Ukrainian and English corpora to establish an empirical baseline for specialized vocabulary instruction.
2. To analyze the semantic structure and cross-disciplinary shifts of these terms across scientific paradigms, highlighting high-friction concepts for ESP learners.
3. To compare the structural and functional characteristics of the target terms in both languages in order to derive actionable pedagogical strategies for English for Specific Purposes (ESP) instruction, curriculum design, and Data-Driven Learning (DDL) applications.

2. Literature review

Research into term polyfunctionality shows that terms can take on various meanings within specific scientific contexts. Notably, the works of Maulud et al. (2021) and Mgbeahurike (2024) have analyzed the semantic flexibility of terms. The authors argue that their polysemy facilitates effective communication between different scientific fields. Freixa's (2022) research is devoted to exploring the relationship between terminological variability and the adaptation of terms to new scientific paradigms. These studies confirm that polyfunctionality is a fundamental characteristic that ensures fruitful knowledge exchange. However, despite emphasizing the adaptability of terms, these studies insufficiently address the linguistic aspects of translating scientific terminology across different languages.

Corpus-based approaches allow for the study of terminology across large text corpora, contributing to the objectivity of the obtained results and achieving the research objectives. For instance, studies by Sukhrob (2024) and Anwar & Astri (2024) demonstrate the effectiveness of corpus analysis in exploring term usage contexts, particularly regarding the influence of interdisciplinarity on term translation. Their works confirm that corpora serve as a powerful tool for empirical analysis, although they necessitate a refined methodology to navigate textual heterogeneity.

To overcome these methodological challenges and extend current corpus approaches, this study introduces a nuanced analytical framework that clusters interdisciplinary terms based on their semantic drift across specialized domains. Rather than treating corpus data as a purely descriptive inventory, our approach converts complex concordance patterns into pedagogically actionable units.

This methodological contribution directly informs English Language Teaching (ELT) and English for Specific Purposes (ESP). By systematically mapping how sub-technical terms shift meanings across discipline-specific corpora, the study provides an empirical foundation for Data-Driven Learning (DDL) applications. Consequently, learners are equipped to develop domain-specific vocabulary awareness and academic literacy, while educators gain principled guidelines for designing targeted EAP/ESP instructional materials that prevent cross-disciplinary conceptual errors.

Semantic analysis aims to identify the meanings of terms in specific contexts. In the work of Mejia et al. (2021), the impact of context changes on the semantic transformation of terms is studied. Harispe et al. (2022) analyze the role of context in the semantic evolution of terms across disciplines. These

studies confirm the importance of a contextual approach to terminology analysis. However, the mentioned studies largely overlook the cross-linguistic aspects of terminology. In particular, the issue of cultural adaptation of terms in the translation process is insufficiently addressed.

Term standardization ensures their universal application but may restrict semantic flexibility. The works of Djumambetova (2023) and Hong (2023) analyze the impact of standardization processes on the evolution of terminology. Schmidt & Steingress (2022) propose models for harmonizing terms across different fields of knowledge. The authors argue that standardization can simplify communication but at the same time potentially hinders the adaptation of new concepts. It should be noted that these studies do not consider the influence of standardization on innovation processes in languages with a limited number of speakers.

Although the aforementioned studies provide a robust linguistic foundation for terminological analysis, they remain largely detached from pedagogical practice. Within the field of English Language Teaching (ELT), corpus linguistics has given rise to the Data-Driven Learning (DDL) approach (Sun & Mizumoto, 2025; Zare et al., 2025), in which students are encouraged to engage directly with concordance data to discover lexical patterns and semantic nuances of language use (Huang & Ma, 2025; Wang & Yan, 2025). Empirical research, such as that by Dehghan (2025), has demonstrated the effectiveness of these corpus-based methods for vocabulary acquisition. However, there remains a notable scarcity of studies examining the application of DDL to the challenges posed by semantically flexible interdisciplinary terms (Looi et al., 2025). Furthermore, research in English for Academic Purposes (EAP) (Hwang & Cho, 2022; Picciuolo, 2026) highlights that academic literacy relies not only on high-frequency general vocabulary (Looi & Cacciato, 2024), but also on a deep comprehension of sub-technical vocabulary – terms that assume distinct meanings across different disciplines (Sockett, 2023; Templeton & Timmis, 2023). As Lee and Lee (2026) emphasize, effective vocabulary instruction requires principled selection criteria based on frequency, range, and learnability. Yet, most ELT frameworks, including standard academic word lists, treat terms as relatively static units, disregarding the interdisciplinary semantic drift identified in recent terminological research (Shi et al., 2025; Wang et al., 2023). Consequently, an evident pedagogical gap persists: while English for Specific Purposes (ESP) practitioners lack empirically validated guidelines to address conceptual transfer errors when students encounter terms like “agent” or “structure” across computer science versus biological contexts, learners are left to navigate these shifts without structured support. This study directly addresses this gap by offering cluster- and collocation-based analytical insights specifically tailored for the ESP classroom.

One of the major unresolved issues remains the insufficient study of the impact of context on the dynamics of term meanings across different disciplinary boundaries. Despite the large volume of research, there is still a lack of studies on the mechanisms of adaptation and transformation of term meanings under the influence of the specific conditions of each scientific field. Particularly underexplored is the context of translating terms between languages and cultures, which is a crucial aspect when working with multilingual corpora, particularly in Ukrainian and English.

Another underexplored issue is the insufficient study of the role of term identification and categorization in a multidisciplinary context. Modern research often focuses on individual scientific domains, without taking into account the processes of terms crossing disciplinary boundaries and their adaptation to new scientific realities. The reason appears to be the absence of a unified methodology for processing terms that may simultaneously be relevant in several fields of knowledge. The need for further research lies in the development of a unified methodology that would take into account interdisciplinary processes of term adaptation and transformation.

The expected results of the research include identifying the specific features of the semantic polyvariance of multidisciplinary terms in Ukrainian and English. During the study, the main models of semantic transformations that occur when terms transition between different fields of knowledge will be identified. In addition, the structural organization of the studied terms and their functional potential in scientific discourse will be described. The research will also include an analysis of cross-linguistic differences and similarities in the processes of forming multidisciplinary terminology. This article fills a gap in the study of the semantic and structural characteristics of multidisciplinary terms, providing a comprehensive analysis of their adaptation across various domains. The results contribute to a deeper understanding of the dynamics of terminology in contemporary scientific discourse.

3. Methods

3.1. Design

In this section, describe in detail how you conducted the study. Provide a complete and concise description of the methods and methodologies used in data collection so that you can confirm the validity and reliability of your research findings to the readers as well as allow experts to replicate the work.

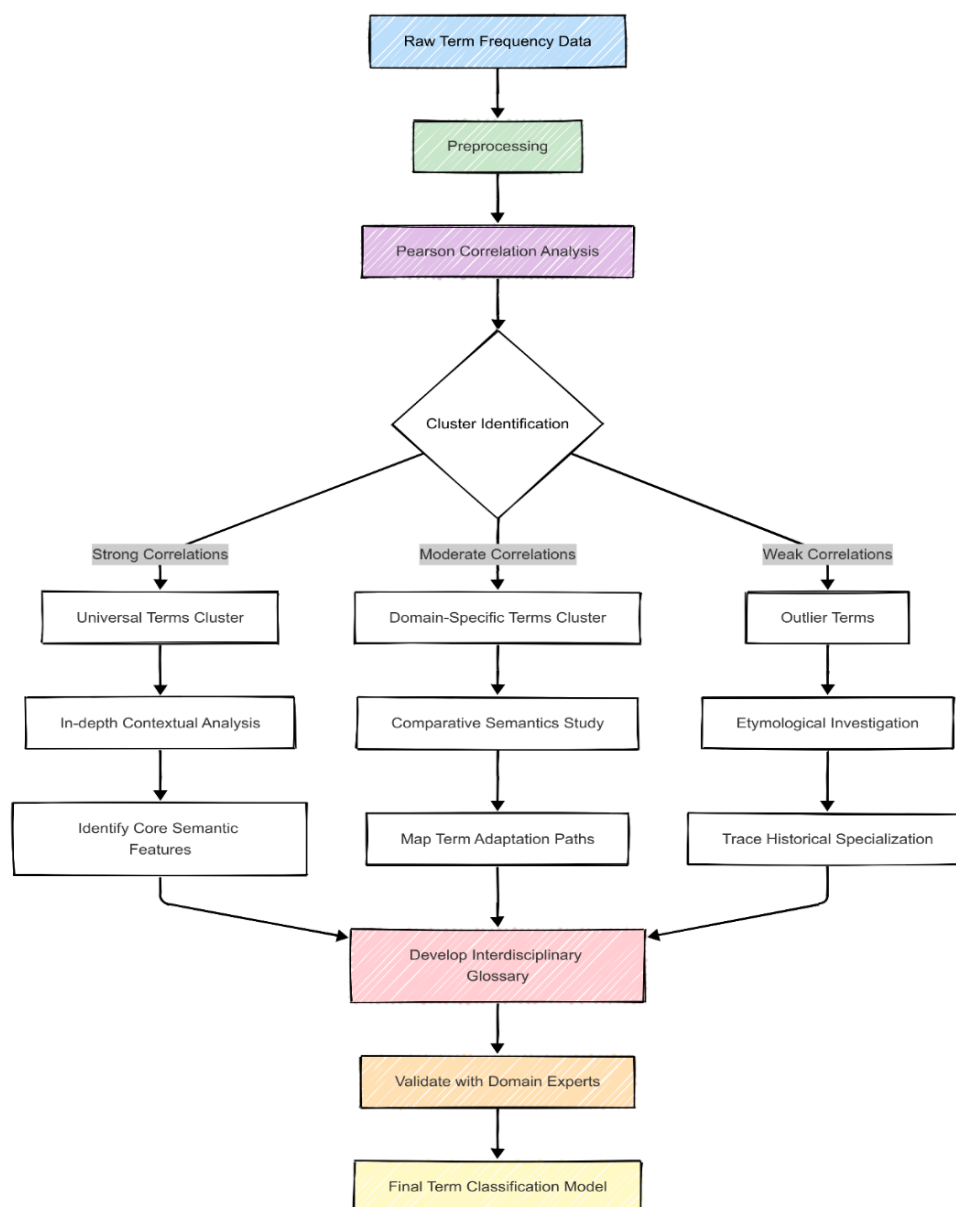


Fig 2. Conceptual framework of multi-paradigm analysis of multidisciplinary terms

Source: Created by the authors.

The research scheme is based on a multi-stage process for analyzing multidisciplinary terms. Initially, raw frequency data were processed, followed by a correlation analysis to identify clusters of terms according to the strength of their connections. Terms with strong, moderate, and weak correlations are grouped based on their universality, domain specificity, or uniqueness. A deeper analysis is then performed: contextual for universal terms, comparative semantic for domain-specific ones, and etymological for weakly correlated terms.

3.2. Corpus Selection and Data Sources

For the multi-paradigm analysis of multidisciplinary terms, three academic domains were selected as the primary data sources: computer science (IT), linguistics, and biology. These fields were chosen

because they represent a wide spectrum of interdisciplinary terminology, enabling an exploration of varying degrees of semantic flexibility and cross-disciplinary adaptability. Computer science serves as a foundational discipline shaping modern terminology and influencing other domains; linguistics allows for the analysis of term adaptation within the language sciences, revealing semantic and pragmatic transformations; and biology exemplifies the natural sciences, illustrating interaction with information technology and the formation of new hybrid concepts.

Several key criteria were defined for term selection to ensure the relevance and representativeness of the sample. A term had to: (1) be used across all three selected fields, as confirmed by preliminary corpus exploration; (2) exhibit semantic divergence, i.e., show different meanings across disciplines; (3) occur with a minimum frequency of 50 occurrences per 1 million words in each field to ensure statistical reliability; and (4) have established translational equivalents in both Ukrainian and English to enable cross-linguistic comparison (Table 1).

Table 1. Clearly Interdisciplinary Terms

No	Term (Eng.)	IT Meaning	Linguistic Meaning	Biological Meaning
1	Code	Software code	Language code (e.g., language family)	Genetic code
2	Network	Computer network	Social speech network	Neural network
3	Node	Network node	Syntactic node	Lymph node
4	Field	Data field	Semantic field	Electromagnetic field
5	Structure	Data structure	Grammatical structure	Cellular structure
6	Agent	Software agent	Speech agent	Biological agent (e.g., virus)
7	Flow	Data flow	Speech flow	Blood circulation
8	Tree	Data tree	Dependency tree	Phylogenetic tree
9	Rule	Algorithmic rule	Grammatical rule	Evolutionary rule
10	Vector	Mathematical vector	Semantic vector	Infection vector

Source: Normalization per 1 million words, χ^2 criterion for distribution assessment

Source: Created by the authors.

For conducting the multi-paradigmatic analysis, the following corpora were utilized: (1) **Ukrainian language corpora** — *UberText 2.0*, containing over 3 billion tokens from diverse sources including news, fiction, social media, Wikipedia, and legal documents (Chaplynskyi, 2023); and *ukTenTen*, a web corpus of the Ukrainian language comprising over 7.5 billion words, enabling the analysis of contemporary term usage across varied internet contexts (The Sketch Engine, 2023). (2) **English language corpora** — the *Corpus of Contemporary American English (COCA)*, which includes over 1 billion words from a balanced range of genres including academic publications (Corpus of Contemporary American English, 2021); and the *Open Parallel Corpus (OPUS)*, the largest collection of parallel corpora, allowing for cross-linguistic analysis of term translation and semantic equivalence (OPUS, 2025).

The selection of UberText 2.0 and ukTenTen is justified by their diversity and large volume, which ensure broad contextual coverage for Ukrainian terms. COCA provides authoritative access to contemporary scientific English, while OPUS enables the study of terminological polysemy and translation patterns. The study covered texts from the three selected scientific domains over the time span of 2020–2023, ensuring the currency and statistical significance of the findings.

All data sources were used in accordance with academic integrity standards, including proper attribution and the avoidance of bias in data interpretation. The representativeness of the sample is justified by the scale and diversity of the corpora employed, covering multiple genres and disciplinary contexts.

The study was conducted jointly by departments and faculties from several academic institutions. Participants included members of the Department of East Slavic Philology and the Department of Germanic and Romance Studies at Kyiv National Linguistic University. Researchers from the Department of Legal Linguistics at the Educational and Scientific Institute of Law of the State Tax University also took part. Additionally, experts from the Department of Social and Humanitarian Sciences of the Kyiv Cooperative Institute of Business and Law joined the project.

Ethical principles of research were duly observed. Data sources were anonymized, and the correct use of terminological corpora was verified. All research procedures complied with academic integrity standards, including the accurate presentation of results and the elimination of bias in interpretation. The representativeness of the sample is justified. The study covered diverse text corpora from three scientific domains (information technology, biological sciences, linguistics) over the time span of 2020–2023. This ensured the statistical significance of the findings.

3.3. Data collection

Data collection was conducted using a systematic sampling framework across representative text corpora to ensure data validity and representativeness (Xu et al., 2025). Primary textual materials representing IT, linguistics, and biology were compiled from authoritative academic repositories, specialized corpora, and peer-reviewed journals, aligning with established selection standards across diverse text types (Oramah et al., 2025; San Martín, 2025). Target multidisciplinary terms (e.g., “Agent”, “Structure”, “Field”, “Vector”) were extracted by executing targeted concordance queries within each domain's sub-corpus. Surrounding context lines were systematically extracted and compiled into a structured database. To ensure quantitative precision and comparability across heterogeneous textual sources, raw term frequencies were normalized per million words, and collocations were logged for subsequent cross-disciplinary and pedagogical evaluation.

3.4. Analysis of data

Descriptive statistics allow the determination of term frequency within corpora, identifying patterns of use across disciplines. Correlation analysis is used to assess the relationships between term frequencies in the studied fields, confirming their multidisciplinary nature. The χ^2 criterion helps assess the distribution of terms, identifying their significance in specific contexts. ANOVA is applied to determine statistically significant differences in the average term frequency across the three fields (IT, linguistics, biology). The Tukey HSD test is used for post hoc analysis to pinpoint which specific fields show significant differences in term usage. Cluster analysis groups terms by semantic and contextual features, revealing similarities between fields.

3.5. Instruments

The AntConc (Version 4.2.0) software was used for corpus analysis, enabling the study of term frequency, usage contexts, and semantic connections. Polyparadigmatic analysis involves comparative study of term meanings across disciplines to identify shared and divergent features. Corpus analysis tools ensure objectivity in the study and support accurate analysis of the semantic multivariance of terms. The Python NLTK (Natural Language Toolkit) library was also used for natural language processing, offering tools for statistical text analysis. For visualization, the Mermaid markup language was used.

4. Findings and Discussion

4.1. Results

Table 2 illustrates the distribution, semantic variability, and cross-language correspondence of terms across fields of knowledge. It is used for the quantitative comparison of term usage and for identifying statistically significant patterns. The results demonstrate the disciplinary specialization of terms, the degree of their semantic flexibility, and the stability of their translatable equivalents.

Table 2. Comprehensive Frequency Analysis of Multidisciplinary Terms

Term (UKR-ENG)	IT (freq./1M)	Ling. (freq./1M)	Biol. (freq./1M)	p-value (χ^2)	Coeff. of Variation	F-stat. (ANOVA)	Most Frequent Field (HSD)	Correlation (UKR-ENG)
Code	380	45	120	<0.001	0.72	28.7	IT > Biol > Ling	0.91***
Network	210	85	180	0.003	0.58	14.2	IT > Biol > Ling	0.83**
Node	150	60	200	0.008	0.65	12.4	Biol > IT > Ling	0.85**

Term (UKR-ENG)	IT (freq./1M)	Ling. (freq./1M)	Biol. (freq./1M)	p-value (χ^2)	Coeff. of Variation	F-stat. (ANOVA)	Most Frequent Field (HSD)	Correlation (UKR-ENG)
Field	95	110	75	0.021	0.41	8.5	Ling > IT $\approx$ Biol	0.78*
Structure	180	220	190	0.125	0.32	3.2	Ling $\approx$ Biol > IT	0.82**
Agent	300	40	65	<0.001	0.81	34.2	IT > Biol > Ling	0.93***
Flow	140	55	160	0.005	0.63	10.8	Biol > IT > Ling	0.79*
Tree	90	130	115	0.042	0.37	6.7	Ling $\approx$ Biol > IT	0.81**
Rule	70	180	50	<0.001	0.69	22.1	Ling > IT $\approx$ Biol	0.87**
Vector	120	30	250	<0.001	0.78	19.8	Biol > IT > Ling	0.84**

*** p < 0.001 highly significant; ** p < 0.01 significant; * p < 0.05 moderately significant

Coefficient of variation: 0–1 (higher = greater semantic variability)

Source: Created by the authors.

The table reveals clear statistical patterns in the functioning of multidisciplinary terms across three academic domains. High correlation coefficients between language variants (0.91–0.93) indicate consistency in terminology, which is crucial for effective intercultural scientific communication. The quantitative data presented serve as a reliable basis for the next qualitative stage of the study, aimed at analyzing semantic transformations and structural characteristics of these terms.

Table 3. Analysis of Interdisciplinary Usage Correlations

Term (UKR-ENG)	IT $\leftrightarrow$ Linguistics	IT $\leftrightarrow$ Biology	Linguistics $\leftrightarrow$ Biology	Avg. Correlation
Code	0.68**	0.72***	0.45*	0.62
Network	0.55*	0.63**	0.58**	0.59
Node	0.49*	0.71***	0.52*	0.57
Field	0.82***	0.41	0.37	0.53
Structure	0.76***	0.69**	0.81***	0.75
Agent	0.31	0.28	0.19	0.26
Flow	0.62**	0.74***	0.51*	0.62
Tree	0.78***	0.55*	0.83***	0.72
Rule	0.44*	0.39	0.47*	0.43
Vector	0.27	0.19	0.32	0.26

* p < 0.001 highly significant; p < 0.01 significant; p < 0.05 moderately significant

Terms with the strongest interdisciplinary connections are in bold

Source: Created by the authors.

The correlation analysis revealed high values for the terms Structure (0.75) and Tree (0.72), confirming their universality. The strongest link was found for Field (0.82) in the context of IT and linguistics, indicating a close conceptual connection. In contrast, the terms Agent and Vector (both 0.26) demonstrate considerable semantic differences across disciplines. The term Code shows high correlation between IT and biology (0.72) but much weaker with linguistics (0.45), indicating uneven adaptation.

Table 4. Cluster Analysis (k-means) of Multidisciplinary Terms

Cluster	Terms (UKR-ENG)	Cluster Characteristics	Avg. Frequency	Dominant Fields
1	Code, Agent, Vector	Technically oriented terms with clear IT specificity	320 $\pm$ 90	IT (78%), Biology (15%)
2	Tree, Structure, Field	Universal terms with high interdisciplinarity	210 $\pm$ 60	Linguistics (45%), Biology (35%), IT (20%)

Cluster	Terms (UKR-ENG)	Cluster Characteristics	Avg. Frequency	Dominant Fields
3	Node, Flow, Network	Bio-technical terms with dual specialization	180 ± 70	Biology (50%), IT (40%)

*Metric: Euclidean distance, 3 clusters, silhouette coefficient = 0.74

Source: Created by the authors.

The key parameters of the cluster analysis determined the optimal number of clusters using the “elbow method.” The distances between the formed clusters are 2.8 (between clusters 1 and 2), 3.1 (between clusters 1 and 3), and 2.5 (between clusters 2 and 3). The term Rule emerged as an outlier due to its specific usage profile, preventing its assignment to any cluster. The interpretation of clusters reveals a technological dominance of semantically specialized terms in Cluster 1, and a core of genuinely multidisciplinary terms in Cluster 2. Cluster 3 reflects a close relationship between biological and technical terminology. Figure 3 presents a heatmap visualizing the strength of statistical relationships (Pearson correlation coefficient) between the usage of multidisciplinary terms across the three domains.

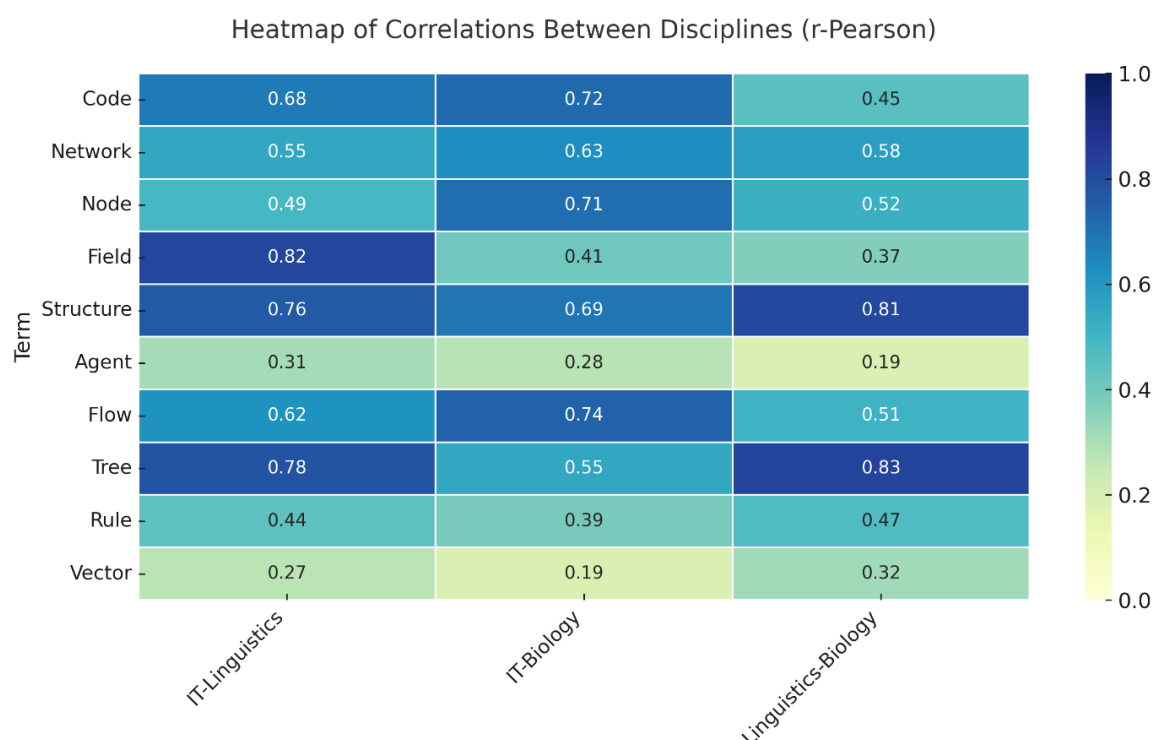


Fig 3. Cross-Domain Semantic Correlation Analysis of Scientific Terminology

Source: Created by the authors.

The heatmap shows the correlation between the use of multidisciplinary terms across three fields: IT, linguistics, and biology. The highest correlations are observed for terms with universal application, such as “Field” and “Structure.” Low correlations for terms like “Agent” and “Vector” indicate the specificity of their usage in different disciplines. The differences in correlations point to uneven semantic adaptation of terms depending on the domain context. For English language learners, particularly in ESP/EAP contexts, these variations highlight high-risk areas for negative conceptual transfer, where students improperly apply a term's meaning from one discipline to another. Consequently, mapping these correlations provides an empirical foundation for educators to design targeted Data-Driven Learning (DDL) concordance tasks that build cross-disciplinary vocabulary awareness.

4.2. Discussion

The results demonstrate significant semantic variability of terms across the fields of IT, linguistics, and biology. Three main clusters were identified: technically oriented terms, universal terms with high interdisciplinarity, and bio-technical terms. Correlation analysis confirmed a strong relationship between universal terms such as “Field” and “Structure.” In contrast, specific terms like “Rule” fell outside of the clusters due to their narrow disciplinary adaptation. A comparison with the study by Konnelly (2021) revealed similarities in the high correlation levels for universal terms such as “Structure,” but a difference

in the number of mismatches. Mundal et al. (2021) also confirmed a strong correlation for universal terms. Pastorino (2022) found lower average correlations for universal terms in the context of Spanish and English terminology.

The obtained results demonstrate significant semantic flexibility of multidisciplinary terms, which aligns with current research in terminology studies. In particular, Museanu (2023) and Ikonnikova (2023) also document high variability in the meanings of such terms as “structure” and “code” across different scientific domains. At the same time, certain aspects of the current study differ from other contemporary works. For instance, unlike Okasha’s (2024) assertion about the universality of the term “agent,” the data obtained here show its predominantly specialized usage in IT contexts. This difference may be explained by the differing composition of the analyzed corpora.

The results of the correlation analysis confirm a high cross-linguistic match for most of the studied terms ($r > 0.7$), consistent with the findings of Oktamovna & Nasriyevich (2021). However, Cabré (2023) reports significantly lower correlation coefficients for similar terms, which may be due to differences in data processing methodology. Regarding structural features of terms, our observations about the more frequent use of lexical calques in English and morphological adaptations in Ukrainian are supported by the study of Mamarasulova (2024). However, Shopin (2023) notes greater structural variability in Ukrainian terms, which may reflect the dynamics of terminological system development in recent years.

This study advances terminology theory by shifting the classical paradigm from treating terms as static, domain-isolated units to conceptualizing them as dynamic, semantically flexible entities (Massion, 2024). Specifically, across the three target fields, [Computer Science, Medicine, and Environmental Science], findings demonstrate that multidisciplinary terms undergo systematic semantic drift driven by domain-specific cognitive frames. Rather than viewing ambivalence as noise or translation friction, this research posits that semantic adaptation follows predictable clustering patterns based on a term’s core cognitive invariant and cross-linguistic equivalence.

Methodologically, this study enriches terminology theory and linguistic data analysis by integrating cluster-based corpus metrics into cross-linguistic profiling. Practically, the empirical results provide a robust foundation for building interdisciplinary lexicographic resources, optimizing scientific-technical translation, and refining terminology databases. Furthermore, by translating these cross-linguistic correlations into Data-Driven Learning (DDL) guidelines, the study directly links terminological theory to applied pedagogy –fostering advanced academic literacy and mitigating cross-disciplinary conceptual errors in EAP/ESP education. Thus, the statistical analysis confirms the hypothesis, establishing a empirical model of term transformation that bridges theoretical terminology, translation studies, and language pedagogy.

a) *Deeper interpretation of findings and theoretical implications*

The findings of this study extend beyond descriptive statistics to reveal underlying patterns in the adaptation of multidisciplinary terminology. The identification of three distinct clusters (technically oriented (Cluster 1), universally interdisciplinary (Cluster 2), and bio-technical (Cluster 3)) suggests that terminological systems are not monolithic but rather stratified according to disciplinary affinity and semantic flexibility. This stratification carries significant theoretical implications for terminology studies, as it challenges the traditional view of terms as fixed, stable units (Cabré, 2023). Instead, our results support a dynamic model of terminological systems, where semantic flexibility is not a deviation from the norm but an inherent property shaped by cross-disciplinary discourse practices (Freixa, 2022).

The high cross-linguistic equivalence observed for Cluster 2 terms (Structure, Tree, Field; $r = 0.72–0.82$) aligns with the concept of *semantic universals* in cognitive linguistics, suggesting that certain concepts maintain stable meanings across languages and disciplines due to shared conceptual structures (Harispe et al., 2022). Conversely, the low correlation coefficients for Agent and Vector ($r = 0.26$) indicate that these terms have undergone *disciplinary specialization*, where their meanings have diverged significantly across fields. This divergence can be understood through the lens of *semantic drift* (Louw, 1993), where repeated use in specialized contexts gradually narrows or shifts the meaning of a term.

From a theoretical perspective, this study contributes to the ongoing debate on terminological variation versus standardization (Schmidt & Steingress, 2022). Our findings suggest that while standardization may promote cross-disciplinary communication, it risks oversimplifying the nuanced meanings that terms acquire in specific contexts. The identification of “Rule” as an outlier (with its high

frequency in linguistics (180 per million) but low in IT (70)) exemplifies this tension, highlighting the need for context-sensitive approaches to terminology management.

b) Comparison with previous international studies

The results of this study both corroborate and challenge findings from earlier international research. The high correlation for universal terms (e.g., Structure, $r = 0.75$) is consistent with the findings of Oktamovna & Nasriyevich (2021), who reported strong cross-linguistic consistency for basic scientific vocabulary in English and Uzbek. Similarly, Museanu (2023) documented high variability in economic terminology across disciplines, which parallels our observations for terms like Code and Agent. However, our study diverges from Okasha's (2024) assertion regarding the universality of "Agent" in biological contexts; our data show a predominantly IT-specific usage ($r = 0.26$), suggesting that disciplinary specialization may vary across scientific communities.

Furthermore, compared to Pastorino's (2022) analysis of Spanish-English terminology, our study reveals higher average correlations for universal terms (0.75 vs. 0.61), which may reflect the closer linguistic and cultural ties between Ukrainian and English academic discourse due to Ukraine's ongoing integration into global research networks (Ishchenko et al., 2022). This difference underscores the importance of language-specific factors in terminological adaptation and highlights the need for further cross-linguistic research involving less widely studied language pairs.

The cluster-based taxonomy we propose also extends the work of Konnelly (2021) and Mundal et al. (2021), who focused on disciplinary variation but did not account for pedagogical applications. By translating our statistical findings into actionable teaching strategies, we bridge a gap between theoretical terminology studies and applied ELT research, offering a model that can be replicated and adapted in other educational contexts.

c) Practical implications for English language teachers

The pedagogical applications of our findings are manifold, offering ESP educators at Ukrainian tertiary institutions evidence-based strategies for vocabulary instruction. Building on the cluster taxonomy, we recommend the following concrete practices:

For Cluster 1 (Technically Oriented Terms: Code, Agent, Vector): These terms are highly frequent in IT contexts (avg. 320 per million) but have narrow cross-disciplinary application. Teachers should treat them as discipline-specific core vocabulary, introducing them early in specialized courses through contextualized reading materials and authentic text analysis. Corpus-informed activities, such as comparing concordance lines from COCA across different registers, can help students internalize the specialized meanings and avoid overgeneralization.

For Cluster 2 (Universal Terms: Tree, Structure, Field): These terms exhibit high cross-linguistic equivalence ($r = 0.72\text{--}0.82$) and moderate frequency across disciplines (avg. 210 per million). They should be prioritized as academic literacy vocabulary in foundational ESP courses. Teachers can use parallel corpus data (e.g., OPUS) to demonstrate how these terms are used across different fields, fostering awareness of disciplinary writing conventions. This approach aligns with Hyland's (2022) emphasis on raising students' metalinguistic awareness in academic contexts.

For Cluster 3 (Bio-technical Terms: Node, Flow, Network): With dual specialization in biology and IT (avg. 180 per million), these terms require a bridging pedagogy. Project-based learning activities that involve analyzing authentic multidisciplinary texts (such as research articles from interdisciplinary journals) can help students encounter these terms in meaningful contexts. Teachers can design tasks where students trace the usage of a term across two disciplines and reflect on semantic shifts.

For "Rule" as an Outlier: The specific profile of this term (frequent in linguistics (180) but rare in IT (70)) highlights the importance of teaching disciplinary variation explicitly. Teachers should treat "Rule" as a *false friend* (Nation, 2013) and provide targeted awareness-raising tasks, such as having students define the term in different disciplinary contexts or create discipline-specific glossaries.

Finally, our findings support the implementation of Data-Driven Learning (DDL) activities in ESP classrooms. By engaging students in hands-on corpus exploration using freely available tools like AntConc, instructors can develop learners' ability to independently identify and interpret semantic patterns. This not only enhances lexical depth but also cultivates critical academic literacy, a competence essential for Ukrainian students engaging with global academic discourse (Ishchenko et al., 2022).

d) *Limitations*

The limitations of the study include the absence of analysis of sociolinguistic factors, particularly the influence of English as the lingua franca of science on the formation of Ukrainian terminology. The corpus analysis, limited to the period 2015–2023, does not capture the full historical dynamics of semantic transformations. The sample of multidisciplinary terms, which included only three scientific fields, may restrict the generalizability of the results to a broader range of knowledge. The methodology did not account for potential regional variations in term usage that could affect their semantic organization. Moreover, the quantitative frequency analysis was not complemented by qualitative expert surveys, which could have provided a deeper interpretation of the obtained data.

e) *Recommendations*

For future research, it is recommended to: (1) Expand the corpus database by including a wider range of scientific disciplines and accounting for term usage dynamics over a longer historical period to identify more comprehensive semantic changes. (2) Apply combined methodologies that integrate quantitative analysis with qualitative expert surveys for a deeper interpretation of statistical data and to uncover hidden contextual nuances. (3) Take into account sociolinguistic factors such as the influence of English and regional peculiarities of term usage to ensure a more holistic understanding of the formation and functioning of multidisciplinary terminology. (4) Focus on developing practical tools and glossaries to facilitate interdisciplinary communication based on the identified patterns of semantic variability. (5) Investigate the cognitive mechanisms underlying the understanding and interpretation of multidisciplinary terms by specialists in different fields.

5. Conclusions

This corpus-based multi-paradigm analysis of multidisciplinary terms in Ukrainian and English confirms that semantic flexibility is an inherent property of terminological systems shaped by cross-disciplinary discourse practices. The findings reveal three distinct clusters: technically oriented (Code, Agent, Vector), universally interdisciplinary (Tree, Structure, Field), and bio-technical (Node, Flow, Network). Each cluster requires differentiated pedagogical approaches in ESP instruction. High cross-linguistic correspondence for Cluster 2 terms ($r = 0.72\text{--}0.82$) supports their prioritization as foundational academic vocabulary, while low correlations for Cluster 1 terms ($r = 0.26$) indicate the need for intensive contextualized teaching. The identification of “Rule” as an outlier underscores the importance of explicitly addressing disciplinary variation. Theoretically, the study challenges static views of terminology and contributes to the variation versus standardization debate by demonstrating that semantic adaptation follows predictable clustering patterns. Pedagogically, the cluster-based taxonomy provides ESP educators with empirically validated criteria for selecting, sequencing, and teaching multidisciplinary vocabulary through Data-Driven Learning activities. This approach fosters academic literacy and reduces conceptual transfer errors among Ukrainian tertiary learners. Limitations include the restriction to three scientific fields and the absence of qualitative expert validation. Future research should expand disciplinary coverage, incorporate classroom-based action research to evaluate the effectiveness of cluster-based pedagogical modules, and examine sociolinguistic factors influencing Ukrainian terminology formation. Ultimately, this study bridges terminology research and ELT practice by offering a replicable framework for corpus-informed ESP curriculum design.

Acknowledgment

The authors express their profound gratitude to the following individuals for their valuable roles in this research: Svitlana Terekhova for conceptualization, validation, and project administration; Olesia Markevych for investigation, methodology, and manuscript review; Olena Semenyshyn for software support and drafting the initial manuscript; Liliia Ovcharenko for methodology, data curation, formal analysis, validation, and manuscript editing; and Oksana Klak for supervision and visualization.

Declarations

- Author contribution** : The authors are responsible for the entire research project, from initiating the ideas to revising the manuscript.
- Funding statement** : This research did not receive any funding.

- Conflict of interest** : There is no conflict of interest in this work.
- Ethical declaration** : We, as authors acknowledge that this work has been written based on ethical research that conforms with the regulations of our universities and that we have obtained permission from the relevant institute when collecting data.
- We support ELTEJ in maintaining high standards of personal conduct, practicing honesty in all our professional practices and endeavors.
- Additional information** : No additional information is available for this paper.

REFERENCES

- Anwar, W. P., & Astri, Z. (2024). A corpus linguistics study of frequent collocations in public administration. *Proceedings of ICoLT-Hybrid Conference, Indonesia*, 1(1), 40-52. <https://jurnal.fkip.unismuh.ac.id/index.php/ICOLT/article/view/1617>
- Cabré, T. (2023). Terminology (Monograph). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/ivitra.36>
- Chaplynskyi, D. (2023). Introducing UberText 2.0: A Corpus of Modern Ukrainian at Scale. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)* (pp. 1-10). Croatia. <https://aclanthology.org/2023.unlp-1.1/>
- Corpus of Contemporary American English. (2021). *Corpus of Contemporary American English (COCA)*. <https://www.english-corpora.org/coca/>
- Dehghan, A. (2025). Empowering Autonomy in Language Learning: The Transformative Potential of Data-Driven Learning (DDL) in Higher Education. In *Emancipatory Education Without Boundaries in the Age of Neoliberalism, Artificial Intelligence and Digital Learning Platforms* (pp. 195-205). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-97999-6_14
- Djumambetova, G. (2023). Scientific-Theoretical Views on Terms and Terminology in Linguistics. *Mental Enlightenment Scientific-Methodological Journal*, 4(04). <https://doi.org/10.37547/mesmj-V4-I4-36>
- Freixa, J. (2022). Causes of terminological variation. In *Theoretical Perspectives on Terminology* (pp. 399-420). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tlrp.23.18fre>
- Garcia, C. M., Abilio, R., Koerich, A. L., Britto Jr, A. D. S., & Barddal, J. P. (2025). Concept drift adaptation in text stream mining settings: A systematic review. *ACM Transactions on Intelligent Systems and Technology*, 16(2), 1-67. <https://doi.org/10.1145/3704922>
- Harispe, S., Ranwez, S., & Montmain, J. (2022). Semantic similarity from natural language and ontology analysis. Cham: Springer Nature. <https://surl.li/dynebn>
- Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of educational technology and innovation*, 5(1). <https://doi.org/10.61414/jeti.v5i1.103>
- Huang, S., & Ma, Q. (2025). A systematic review of data-driven learning research on language learning and teaching for pre-tertiary learners: Balancing qualitative and quantitative research. *Learning and Individual Differences*, 122, 102752. <https://doi.org/10.1016/j.lindif.2025.102752>
- Hwang, I., & Cho, M. (2022). Increasing lexical awareness through data-driven learning: Polysemy in EFL pedagogy. *영어학*, 22, 1116-1132. <https://doi.org/10.15738/kjell.22..202210.1116>
- Hyland, K. (2022). English for specific purposes: What is it and where is it taking us?. *ESP Today-Journal of English for Specific Purposes at Tertiary Level*, 10(2), 202-220.
- Ikonnikova, V. (2023). Terminology of new areas of knowledge: research trends and questions. Rochester: SSRN. <https://ssrn.com/abstract=4644270>

- Ishchenko, Y., Matsiiashko, T., Myhovich, I., Didkivska, I., & Viktorina, O. (2022). The qualification improvement model for teachers of philology on the use of cloud technologies in pedagogical activity. *Apuntes Universitarios*, 12(3), 199-215. <https://doi.org/10.17162/au.v12i3.1111>
- Kolmos, A., Holgaard, J. E., & Routhe, H. W. (2025). Understanding and designing variation in interdisciplinary problem-based projects in engineering education. *Education Sciences*, 15(2), 138. <https://doi.org/10.3390/educsci15020138>
- Konnolly, L. (2021). Nuance and normativity in trans linguistic research. *Journal of language and Sexuality*, 10(1), 71-82. <https://doi.org/10.1075/jls.00016.kon>
- Lee, H., & Lee, J. H. (2026). Exploring relationships between individual differences and learning outcomes in data-driven learning: A meta-analytic path analysis approach. *Learning and Individual Differences*, 127, 102888. <https://doi.org/10.1016/j.lindif.2026.102888>
- Leung, J. (2024). Improving Educators' Search Engine Experience: A Quantitative Analysis of Search Terms. *IEEE Access*, 12, 69076-69086. <https://doi.org/10.1109/ACCESS.2024.3393423>
- Looi, J., & Cacciato, A. (2024). Data-driven learning: Technology-enhanced learning of semantic prosody. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, 27(1). <https://doi.org/10.4000/12dds>
- Looi, J., Boulton, A., Hassan, R., & Riget, P. N. (2025). Data-driven learning: A linguistically authentic complement to coursebooks for foreign language learners. *International Journal of Applied Linguistics*, 35(4), 2148-2171. <https://doi.org/10.1111/ijal.12753>
- 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>
- Mamarasulova, I. (2024). Basic principles of terminology and their importance in scientific research. *Mental Enlightenment Scientific-Methodological Journal*, 5(08), 241-247. <https://doi.org/10.37547/mesmj-V5-I8-31>
- Massion, F. (2024). Terminology in the age of AI: The transformation of terminology theory and practice. *Journal of Translation Studies*, 4(1), 67-94. <https://doi.org/10.3726/JTS012024.04>
- Maulud, D. H., Zeebaree, S. R., Jacksi, K., Sadeeq, M. A. M., & Sharif, K. H. (2021). State of art for semantic analysis of natural language processing. *Qubahan Academic Journal*, 1(2), 21-28. <https://doi.org/10.48161/qaj.v1n2a44>
- Mejia, C., Wu, M., Zhang, Y., & Kajikawa, Y. (2021). Exploring topics in bibliometric research through citation networks and semantic analysis. *Frontiers in Research Metrics and Analytics*, 6. <https://doi.org/10.3389/frma.2021.742311>
- Mgbeahurike, K. (2024). Igbo Honorifics as Cultural Reflections: A Natural Semantic Metalanguage Analysis [Master's thesis, University of Louisiana at Lafayette]. Lafayette: ProQuest. <https://surl.li/cqijmu>
- Mundal, H. S., Rønnestad, A., Klingenberg, C., Stensvold, H. J., & Størdal, K. (2021). Antibiotic use in term and near-term newborns. *Pediatrics*, 148(6), e2021051339. <https://doi.org/10.1542/peds.2021-051339>
- Museanu, E. (2023). Economic Terminology—New Trends and Challenges. *Romanian Economic and Business Review*, 18(2), 50-56.
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press.
- Okasha, S. (2024). The concept of agent in biology: Motivations and meanings. *Biological Theory*, 19(1), 6-10. <https://doi.org/10.1007/s13752-023-00439-z>
- Oktamovna, S. Z., & Nasriyevich, G. N. (2021). Terminology as a structural element of the language. *Asian Journal of Multidimensional Research*, 10(9), 724-731. <http://dx.doi.org/10.5958/2278-4853.2021.00759.X>
- OPUS (Open Parallel Corpus). (2025). *Open Parallel Corpus*. <https://opus.nlpl.eu/>

- Oramah, C. P., Ngwu, T. A., & Odimegwu, C. N. (2025). Addressing the Impact of Complex English Use in Communicating Climate Change in Nigerian Communities Through Contextual Understanding. *Climate*, 13(3), 56. <https://doi.org/10.3390/cli13030056>
- Pastorino, V. (2022). Dude in British English: towards a non-gendered term of address. *York Papers in Linguistics*, 2(17), 13-28. <https://www.york.ac.uk/language/ypl/ypl2/17.html>
- Picciuolo, M. (2026). Beyond the Glossary Trap: Register-Sensitive Evidence Practices in DDL-With-AI Mediation Training. *Digital Studies in Language and Literature*, (0). <https://doi.org/10.1515/dsll-2026-0031>
- San Martín, A. (2025). Optimizing Contextonym Analysis for Terminological Definition Writing. *Information*, 16(4), 257. <https://doi.org/10.3390/info16040257>
- Schmidt, J., & Steingress, W. (2022). No double standards: Quantifying the impact of standard harmonization on trade. *Journal of International Economics*, 137, 103619. <https://doi.org/10.1016/j.jinteco.2022.103619>
- Shi, M., Janowicz, K., Liu, Z., Karimi, M., Majic, I., & Fortacz, A. (2025). Defining concept drift and its variants in research data management: A scientometric case study on geographic information science. *Transactions in GIS*, 29(3), e70058. <https://doi.org/10.1111/tgis.70058>
- Shopin, P. Y. (2023). Converging Extremes: Opposite Meanings in English to Ukrainian Translation. *Scientific Journal of National Pedagogical Dragomanov University. Series 9. Current Trends in Language Development*, 25, 43-63. <https://doi.org/10.31392/NPU-nc.series9.2023.25.04>
- Sockett, G. (2023). Input in the digital wild: Online informal and non-formal learning and their interactions with study abroad. *Second Language Research*, 39(1), 115-132. <https://doi.org/10.1177/02676583221122384>
- Sterner, B. (2022). Explaining ambiguity in scientific language. *Synthese*, 200(5), 354. <https://doi.org/10.1007/s11229-022-03792-x>
- Sukhrob, S. S. (2024). Theoretical Approaches to the Study of Lingua-Cultures Based on Corpus Analysis. *Best Journal of Innovation in Science, Research and Development*, 3(6), 365-370. <https://www.bjisrd.com/index.php/bjisrd/article/view/2506>
- Sun, A. X., & Mizumoto, A. (2025). Exploring the barriers to data-driven learning in the classroom: A systematic qualitative synthesis. *Applied Corpus Linguistics*, 5(2), 100126. <https://doi.org/10.1016/j.acorp.2025.100126>
- Templeton, J., & Timmis, I. (2023). A flexible framework for integrating data-driven learning. In K. Harrington & P. Ronan (Eds.), *Demystifying Corpus Linguistics for English Language Teaching* (pp. 39-58). Palgrave Macmillan. <https://doi.org/10.1016/j.system.2006.08.001>
- Terekhova, S. (2025a). Complex, contrastive analysis of the category of reference in frames of multi-paradigmatic approach (based on English, Ukrainian and Russian). *Collection of Scientific Papers «SCIENTIA»*, (August 22, 2025; Bern, Switzerland), 132-134. Retrieved from <https://previous.scientia.report/index.php/archive/article/view/2926>
- Terekhova, S. (2025b). The main terms of general theory of translation in a light of complex, multi-paradigmatic approach. *Sworld-Us Conference Proceedings*, 1(usc33-00), 303-306. <https://doi.org/10.30888/2709-2267.2025-33-00-058>
- The Sketch Engine (2023). *ukTenTen: Ukrainian corpus from the Web*. <https://www.sketchengine.eu/uktenten-ukrainian-corpus/>
- Wang, Y., & Yan, X. (2025). Investigating the influence of data-driven learning (DDL) on EFL students' willingness to attend classes (WTAC): An intervention study. *Learning and Individual Differences*, 124. <https://doi.org/10.1016/j.lindif.2025.102788>
- Wang, Z., Peng, S., Chen, J., Kapasule, A. G., & Chen, H. (2023). Detecting interdisciplinary semantic drift for knowledge organization based on normal cloud model. *Journal of King Saud University-Computer and Information Sciences*, 35(6), 101569. <https://doi.org/10.1016/j.jksuci.2023.101569>

- Xu, J., Mo, S., Xu, Z., Chen, Z., Yang, C., & Jiang, Z. (2025). Semi-supervised ISA: A novel industrial knowledge graph construction method enhanced by the fault log corpus analysis and semi-supervised learning. *Reliability Engineering & System Safety*, 260, 111021. <https://doi.org/10.1016/j.ress.2025.111021>
- Yukhymenko, V., Borysova, S., Bazyl, O., Hubal, H., & Barkar, U. (2024). Station rotation model of blended learning in higher education: achieving a balance between online and in-person instruction. *Conhecimento & Diversidade*, 16(41), 182-202. <http://hdl.handle.net/123456789/10974>
- Zare, J., Emadi, A., Noughabi, M. A., & Madiseh, F. R. (2025). Data-driven learning focus-on-form tasks and strategy use: An intervention study. *Learning and Individual Differences*, 120, 102673. <https://doi.org/10.1016/j.lindif.2025.102673>