

UDC 378.147

IRSTI 06.81.89

<https://doi.org/10.48371/PEDS.2026.81.2.018>

**METHODOLOGICAL ASPECTS OF USING COMPUTER-BASED
ANALYSIS OF ARCHITECTURAL TERMINOLOGY IN TEACHING
PROFESSIONALLY ORIENTED FOREIGN LANGUAGE**

*Smagulova Zh.A.¹, Zhylytyrova Zh.T.², Kalelova I.M.³

*^{1,2}Ablai Khan University of International Relations and World Languages,
Almaty, Kazakhstan

³East Kazakhstan Technical University named after D.Serikbaev,
Oskemen, Kazakhstan

Abstract. The present article examines the distribution of architectural terminology across English for Specific Purposes (ESP) courses designed for future architects at three proficiency levels: elementary (A1), pre-intermediate (A2), and intermediate (B1). The empirical foundation of the study is a purpose-built corpus comprising texts extracted from three professionally oriented textbooks: *Professional English in Use: Architecture* (Konovalova E.N.), *English for Architects* (Bezruchko E.N.), and *English for Construction Universities* (Lukina L.V.). The corpus texts were digitised and pre-processed through case normalisation, stop-word removal, and tokenisation, then subjected to frequency and topic modelling analysis employing Python, NLTK, and gensim libraries. Frequency distributions and visualisations - bar charts, pie charts, and pyLDAvis topic maps - enabled the identification of a compact high-frequency core of architectural vocabulary (architecture, building, construction, materials, design), three thematic clusters (structural elements, materials, design principles), and one dominant LDA topic representing the conceptual nucleus of the corpus. Drawing on frequency data, topic membership, and contextual usage patterns, a level-differentiated model of terminology distribution for the A1–B1 continuum is proposed. This model informs the design of VR-based learning tasks, ranging from naming and describing basic spatial elements in a virtual environment at A1 to complex project assignments integrating form, style, and sustainability concepts at B1. The findings demonstrate that combining corpus-linguistic methods with immersive VR technologies affords a more empirically grounded trajectory for architectural vocabulary acquisition, aligning lexical selection with authentic frequency profiles and the thematic architecture of academic texts, thereby enhancing the overall effectiveness of ESP instruction.

Keywords: computer-based corpus analysis, architectural terminology, English for Specific Purposes (ESP), frequency analysis, LDA topic modelling, VR-based language learning, structural elements, design principles

Introduction

Computer analysis of architectural texts opens new opportunities for evidence-based selection of terminology in ESP for future architects. Unlike purely intuitive vocabulary selection, corpus and topic methods rely on real usage cycles of terms, their typical contexts and thematic clusters, which is especially important when designing tasks for levels A1, A2 and B1. The relevance of the study is due to the fact that a systematic selection of architectural terminology for different proficiency levels is a key condition for the quality of professionally oriented language courses for architecture students. The empirical base of the study is a corpus compiled from three textbooks: Professional English in Use: Architecture (Konovalova E.N.), English for Architects (Bezruchko E.N.) and English for Construction Universities (Lukina L.V.) [1-3]. These textbooks are actively used in the training of future architects and will later be integrated into VR-based tasks, where students interact with virtual objects and navigate them in English. The aim of the study is to conduct a computer-based analysis of architectural terminology in these textbooks and to create a level-based lexical model (A1–B1) for subsequent use in designing VR tasks in ESP courses. The objectives are: (1) to build a unified text corpus; (2) to carry out pre-processing and frequency analysis; (3) to visualise the distribution of key terms; (4) to identify thematic clusters using LDA; and (5) to propose a level-based terminology model that takes into account VR-based forms of instruction. The scientific significance of the work lies in the use of objective corpus data and methods of computational linguistics to justify course content instead of relying only on the teacher’s intuition. The scientific novelty of the study consists in combining corpus analysis of architectural terminology, its level-based distribution along the A1-B1 scale, and the integration of these results into the design of VR tasks; existing studies usually either describe terminology without corpus support, or do not consider the potential of virtual reality in ESP course design.

Materials and methods

The corpus includes all text materials from the three textbooks, digitised as PDF files and uploaded for analysis. Text was extracted page by page using the PyPDF2 library; then the texts were concatenated into a single file and saved for further processing. The methodological basis of the study includes a corpus-based approach, frequency analysis and LDA topic modelling implemented in Python with PyPDF2, NLTK and gensim, which ensures reproducibility and transparency of all data processing stages [4].

Pre-processing and frequency analysis. The corpus text was converted to lower case; non-letter characters were removed using regular expressions so that only letters and spaces remained, and tokenisation was performed by splitting on spaces. General-language words were removed using NLTK stop-word lists

for English and Russian and a minimum token length of two characters. Word frequencies were calculated with the Counter class; a top-20 list was created, and a set of rare words was collected, among which many architectural terms appeared. Some very frequent general words (th, unit, used, one, text, new, exercise, word, following) were excluded at an additional filtering stage. The results of the frequency analysis were visualised in two ways: a bar chart of the most frequent words after filtering (frequency on the X-axis, lexemes on the Y-axis, labels above the bars) and a pie chart “Frequency distribution of specific words” for nine key terms (architecture, materials, construction, buildings, design, water, work, form, fire) with their absolute frequencies and percentage shares.

LDA topic modelling. To reveal hidden thematic structures in the corpus, an LDA model (LdaMulticore, gensim) with num_topics = 5 was built on the basis of a dictionary and a bag-of-words representation (Fig. 2) [5]. The number of topics was set to five following an iterative evaluation of coherence scores (u_{mass} metric): values of k = 3 through k = 8 were tested, and k = 5 yielded the highest coherence (-2.84) while producing thematically interpretable and non-overlapping clusters; smaller k collapsed distinct domains, whereas larger k introduced redundant micro-topics with low discriminatory value. For each topic, top words were listed, and the results were visualised with pyLDAvis.

Results and discussion

High-frequency terminology. Frequency analysis revealed two main layers of vocabulary: a general academic layer (exercise, text, task, people, system, etc.) and a specialised architectural-construction layer (architecture, building, materials, construction, design, roof, foundation, etc.). The pie chart (Fig. 1) shows that among the selected terms the leaders are architecture (556 occurrences, 19.2%), water (385; 13.3%), materials (315; 10.9%), construction (310; 10.7%) and fire (303; 10.5%), while design (269), buildings (256), work (253) and form (249) have slightly lower but still substantial shares. This distribution confirms the existence of a compact terminological core related to the description of structures and materials and highlights the importance of engineering aspects (water, fire, work) in the teaching texts. For teaching purposes, this core can serve as a starting point when selecting vocabulary for VR scenarios where students describe materials, structural solutions and building systems.

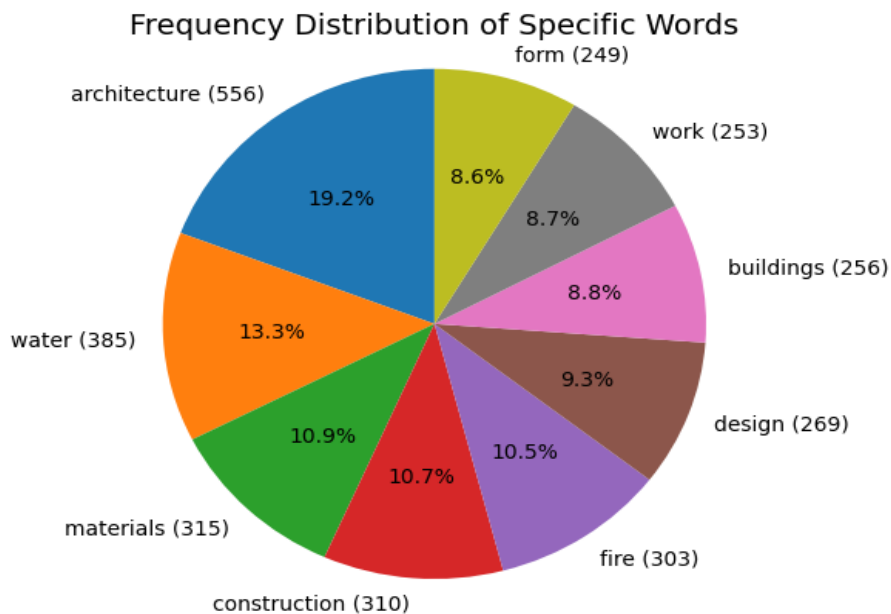


Figure 1 - Frequency Distribution of Specific Words

LDA thematic clusters. The LDA model (Fig. 2) identified five interpretable topics. Topic 1 represents a general academic and engineering context (exercise, system, task, text, people, water, computer, money) related to the description of the learning process and engineering systems. Topic 2 focuses on grammatical material (fire, participle, questions, example, ed, English, etc.), that is, exercises on participles and verb forms; Topic 3 represents the English-language context for architects (English, example, time, architects, site), including examination situations. Topic 4 covers complex grammatical structures (adverbs, verbs, parts, subject, complex), which indicates a substantial block of general language teaching material. Topic 5, which dominates in terms of token share, forms the core of architectural concepts and includes words such as architecture, building, unit, text, following, new, used, style, form, use; in the pyLDavis visualisation this appears as a large cluster with high weight and a bar chart of the most relevant terms [6].

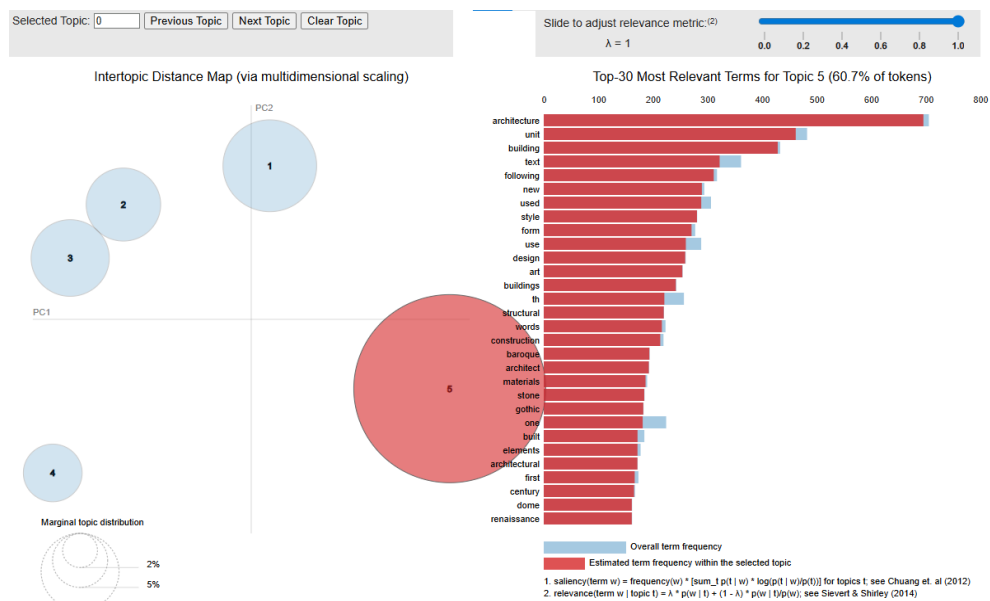


Figure 2 - LDA topic 5: intertopic distance map and top-30 architectural terms

Thematic grouping and A1-B1 levels. On the basis of frequency data and LDA topics, architectural terms were manually grouped into three conceptual categories (Fig. 3): structural elements (arch, beam, column, dome, foundation, lintel, steel, structure, truss, vault, etc.); design principles (design, form, proportion, style, symmetry, etc.); and materials (brick, concrete, glass, steel, timber, etc.). The assignment of terms to CEFR levels followed three explicit criteria applied in sequence. First, corpus frequency: high-frequency terms (appearing ≥ 15 times in the corpus) were assigned to A1, medium-frequency terms (5-14 occurrences) to A2, and low-frequency or multi-word terms (< 5 occurrences) to B1. Second, cognitive complexity: concrete, visually salient referents (e.g., door, column) were placed at A1, functional or process-related concepts (e.g., cantilever, insulation) at A2, and abstract or stylistic concepts (e.g., fenestration, rustication) at B1. Third, VR task affordance: A1 terms support identification and labelling tasks in a virtual environment; A2 terms enable explanation and description tasks; B1 terms underpin complex project and evaluation tasks involving form, style and sustainability discourse.

Architectural Terms Categorized by Theme:

--- Structural Elements ---

- | | | | |
|-----------|-------------|---------------|---------------|
| 1. arch | 2. beam | 3. buttress | 4. cantilever |
| 5. column | 6. concrete | 7. dome | 8. foundation |
| 9. lintel | 10. steel | 11. structure | 12. truss |
| 13. vault | | | |

--- Design Principles ---

- | | | | |
|-----------|-----------|-----------------|----------|
| 1. design | 2. facade | 3. fenestration | 4. space |
|-----------|-----------|-----------------|----------|

--- Materials ---

- | | | | |
|--------------|-------------|----------|---------------|
| 1. brick | 2. concrete | 3. glass | 4. insulation |
| 5. materials | 6. mortar | 7. panel | 8. plaster |
| 9. steel | 10. tile | | |

Figure 3 - Architectural Terms Categorized by Theme

The bar chart “Conceptual distribution of architectural terms by theme” (Fig. 4) shows a predominance of vocabulary related to structural elements (about 13 terms) and materials (10 terms) compared with a smaller group of items describing design principles (4 terms). This pattern corresponds to the practical needs of ESP courses, where students more often describe structures and materials of buildings in real or virtual environments [7].

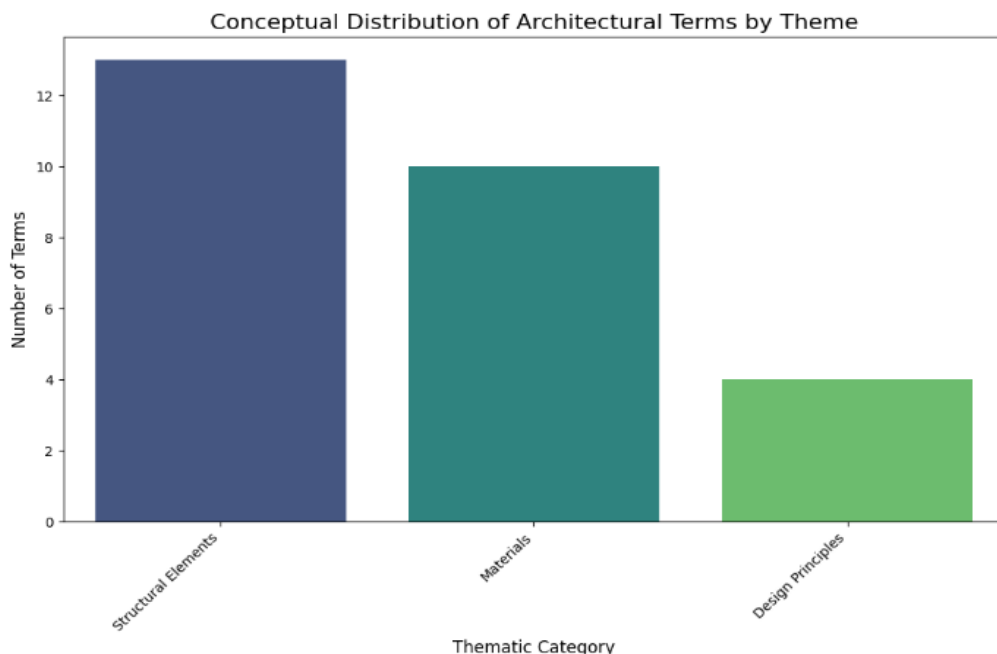


Figure 4 - Conceptual Distribution of Architectural terms by Theme

Taking into account frequency, thematic affiliation and cognitive complexity, a level-based terminology model was proposed (*Fig. 5*).

- Beginner A1: basic terms for naming objects and their parts (arch, beam, brick, building, column, concrete, door, floor, glass).

- Continuing A2: terms related to functional zones and simple engineering aspects (facade, buttress, cantilever, dome, insulation, foundation, structure, truss, vault).

- Continuing B1: more complex and multi-word concepts (atrium, capital, cornice, fenestration, frieze, pilaster).

<i>Level A1</i>	<i>Level A2</i>	<i>Level B1</i>
Начинающий	Средний	Высокий
=====	=====	=====
1. Arch	1. Buttress	1. Atrium
2. Beam	2. Cantilever	2. Capital
3. Brick	3. Dome	3. Cornice
4. Building	4. Facade	4. Fenestration
5. Column	5. Foundation	5. Frieze
6. Concrete	6. Insulation	6. Mezzanine
7. Design	7. Lintel	7. Pediment
8. Door	8. Structure	8. Pilaster
9. Floor	9. Truss	
10. Glass	10. Vault	

Figure 5 - Level-based distribution of architectural terms across A1, A2 and B1

The corpus-based analysis demonstrates that architectural terminology in the three ESP textbooks is strongly structured around a compact lexical core (architecture, building, construction, materials, water, fire, design, work, form). This core reflects not only the physical components of buildings but also key engineering concerns such as safety (fire), durability and performance (water, materials), and functional organisation (work, form). Such a distribution suggests that the textbooks prioritise basic structural and safety-related concepts as the foundation of professional discourse, which is pedagogically justified for beginner and lower-intermediate levels of ESP for future architects. At the same time, the frequency profile indicates a certain imbalance between purely architectural-conceptual terms and broader engineering or environmental notions [8]. The relatively high frequencies of water and fire show that risk, protection and building services are systematically integrated into the teaching material, while more abstract design concepts (form, style, aesthetics, sustainability) occupy a more modest but still significant share of the core vocabulary. This can be interpreted as an orientation of the courses towards practice-oriented, technical competences rather than purely stylistic or theoretical reflection, which is typical for ESP textbooks aimed at future practitioners. The LDA topic modelling results deepen

this picture by revealing stable thematic clusters, such as structural elements, materials and design principles. These topics correlate with the main professional competences of future architects: the ability to identify and describe building parts, to justify material choices, and to discuss design solutions in terms of function, safety and performance. The presence of a dominant topic representing the core of architectural concepts suggests that the corpus has a strong thematic focus, which can be exploited in curriculum design by aligning VR scenarios with the most central conceptual fields rather than dispersing attention across marginal or low-frequency topics [9]. From a level-based perspective (A1-B1), the data allow a more nuanced allocation of terminology than a purely intuitive approach. High-frequency, concrete nouns (e.g. wall, roof, door, window, materials, construction, building) are clearly suitable for A1, where learners need to name and recognise basic elements in the virtual environment and perform simple identification tasks. Less frequent but still central terms related to functions and processes (e.g. foundation, insulation, structure, load, ventilation) can be introduced at A2, when students start describing how buildings work and comparing alternative solutions. More complex and abstract items associated with design principles and sustainability (e.g. form, style, efficiency, sustainability, environment, urban context) are more appropriate for B1, where learners are expected to discuss and justify design choices and reflect on wider architectural concepts. The level-based interpretation of the corpus data can be further clarified by mapping lexical clusters onto proficiency levels and corresponding VR task types. *Table 1* summarises how the three main thematic clusters identified in the analysis - structural elements, materials and properties, and design and sustainability - can be distributed across levels A1-B1 and operationalised in VR activities. This mapping makes explicit the progression from simple naming of building parts to more complex reasoning about material choice and design decisions in immersive tasks.

Table 1. Distribution of architectural terminological clusters across A1-B1 levels and corresponding VR task types

<i>Cluster / focus</i>	<i>Typical terms (examples)</i>	<i>Suggested level</i>	<i>VR task focus</i>
Structural elements	wall, roof, foundation, column, beam	A1	Naming and locating parts of a virtual building
Materials and properties	concrete, steel, glass, insulation, strength	A2	Comparing materials and explaining basic functional roles
Design and sustainability	form, style, façade, efficiency, sustainability	B1	Justifying design choices and discussing sustainable solutions in projects

As shown in *Table 1*, the A1 cluster is dominated by high-frequency, concrete nouns that correspond to visually salient objects in the virtual environment, which supports simple identification, naming and matching tasks. At A2, VR scenarios can shift towards explaining how buildings work by foregrounding material properties and functional aspects (e.g. why a particular material is chosen for a roof or façade), while at B1, tasks can require learners to articulate and defend design and sustainability decisions, using more abstract and argumentative language that reflects the higher-level conceptual vocabulary revealed by the corpus. In this sense, the corpus analysis does not merely list words but provides empirical evidence for sequencing terminology across proficiency levels in a principled way. The transition from A1 to B1 can be seen as a gradual movement from naming static objects to reasoning about dynamic properties, design strategies and environmental implications. Such an evidence-based progression is particularly important for VR-based tasks, where cognitive and linguistic load must be carefully controlled: early tasks can focus on identifying and labelling visible components, while more advanced tasks can require evaluating alternative solutions, explaining trade-offs and integrating sustainability considerations into design discussions. The integration of corpus methods and VR design also has methodological implications for ESP course development. Compared to intuitive vocabulary selection, frequency analysis and topic modelling reduce subjectivity by showing which terms actually dominate the textbooks and how they cluster thematically, thus allowing the course designer to verify whether planned VR scenarios genuinely reflect the lexical reality of the materials used in class. However, the approach has limitations: the corpus is relatively small and based solely on three textbooks, which may not fully represent the language of professional practice or academic research; furthermore, corpus preprocessing (stopword removal, exclusion of very frequent general words, focus on single tokens) may have obscured some useful multiword expressions, collocations and phraseological patterns crucial for professional communication. These limitations suggest several directions for further research and refinement of the model. Future studies could expand the corpus with authentic architectural texts (e.g. professional guidelines, project descriptions, research articles) and spoken data from design studios in order to compare textbook terminology with real professional usage and adjust the VR tasks accordingly. Another promising avenue is a more detailed analysis of collocations and lexical bundles specific to architecture (e.g. load-bearing wall, sustainable materials, fire safety regulations), which can be integrated into VR scenarios that require not only isolated word recognition but also production of more natural, domain-specific phrases [10]. Despite these constraints, the findings support a more systematic, data-driven approach to ESP course design for architecture students. The identified lexical core and thematic clusters can serve as a reference for selecting and sequencing vocabulary, while the level-based model directly informs the complexity and focus

of VR tasks at A1, A2 and B1. In this way, corpus analysis and VR technology complement each other: corpus data provide a solid empirical foundation for what should be taught, and VR offers an immersive environment for how it can be practised in realistic, professionally relevant situations, potentially increasing both the effectiveness and the motivation in ESP instruction for future architects.

Limitations. Several limitations of the present study should be acknowledged. First, the corpus is confined to three textbooks and excludes authentic professional discourse (project briefs, tender documents, architectural reviews); the lexical model therefore reflects textbook language rather than workplace usage. Second, level assignments (A1–B1) relied on expert judgement combined with frequency and coherence criteria; inter-rater reliability was not formally measured and should be established in future research. Third, as an unsupervised probabilistic model, LDA introduces inherent uncertainty in topic boundaries; the five-topic solution represents one plausible interpretation. Finally, the proposed VR task types have not yet been piloted empirically; pedagogical claims remain theoretical and will be validated in the experimental phase of the study.

Notwithstanding these limitations, the findings yield several theoretically and practically significant conclusions. The corpus compiled from three ESP architecture textbooks yields a compact high-frequency lexical core (architecture, building, construction, materials, design, form) that reflects both physical and functional dimensions of professional architectural discourse. Frequency analysis, LDA topic modelling and data visualisation produced consistent, mutually reinforcing results, confirming the viability of the combined computational approach for evidence-based vocabulary selection in ESP instruction. Building on these results, the present study demonstrates, for the first time on an architecture-textbook corpus, a coherent link between corpus analysis, CEFR-level terminology distribution and VR-oriented task design, thereby bridging computational linguistics and pedagogical decision-making in architectural ESP. More specifically, the five-topic LDA solution maps onto three conceptual clusters (structural elements, materials, design principles), providing an empirically grounded basis for the A1-B1 lexical model and the corresponding VR scenario framework.

The computer analysis of the corpus compiled from *Professional English in Use: Architecture* (Konovalova E.N.), *English for Architects* (Bezruchko E.N.) and *English for Construction Universities* (Lukina L.V.) confirmed that combining frequency analysis, LDA topic modelling and data visualisation is an effective way to describe and structure architectural terminology in ESP. In contrast to traditional, intuition-based selection of vocabulary, this approach reveals an empirically grounded lexical core and stable thematic clusters that reflect the real distribution of concepts in the teaching materials, providing a more objective basis for syllabus design.

The study is the first, for a corpus of architecture textbooks, to demonstrate a coherent link between corpus analysis, level-based distribution of terms and VR-oriented task design, thus bridging the gap between computational linguistics and pedagogical decision-making in ESP for architecture. The resulting frequency diagrams and topic maps make it possible not only to identify a compact set of key architectural concepts, but also to map them onto proficiency levels A1, A2 and B1, producing a terminology model that supports a gradual progression from naming basic elements to discussing complex design and sustainability issues. This model directly informs the design of VR activities, where lexical difficulty, thematic density and cognitive load can be calibrated in line with the actual lexical profile of the textbooks rather than arbitrary assumptions. From a broader perspective, the findings suggest that corpus-driven, VR-integrated course design can increase both the validity and the relevance of ESP instruction for architecture students. When VR scenarios are built around empirically identified high-frequency terms and central topics, learners are more likely to practise the language that they will encounter in core disciplinary texts and professional communication, which strengthens the alignment between classroom tasks, textbook content and real-world needs. At the same time, the explicit level-based model makes it easier for teachers and curriculum developers to coordinate lexical progression with other components of competence, such as grammar, skills and professional tasks.

Future work will focus on extending the corpus with architectural design texts, regulatory and normative documents, and, where possible, authentic project descriptions, in order to compare textbook language with professional discourse and adjust the lexical model accordingly. Further refinement of the level-based term lists using international proficiency descriptors (e.g. CEFR-aligned can-do statements) will help to anchor the model in widely recognised standards and to make it transferable to other institutions and contexts. Another important direction is the development and empirical testing of VR scenarios whose frequency and thematic characteristics deliberately mirror the identified lexical core and topic structure, so that tasks for describing, analysing and evaluating architectural objects can be systematically scaffolded across levels and evaluated in terms of learning outcomes.

REFERENCES

- [1] Konovalova E.N. Professional English in Use: Architecture. Textbook for architecture students. – Moscow: Cambridge University Press / MCBA. - 144 p.
- [2] Bezruchko E.N. English for Architects. Coursebook for students of architecture. – Moscow: ASV Publishing House. – 176 p.
- [3] Lukina L.V. English for Construction Universities. ESP textbook for construction and architecture majors. – Moscow: MGSU Publishing House. - 200 p.

[4] Boulton, A. Integrating corpus tools and techniques in ESP courses // *ASp. La revue du GERAS*. – 2016. – № 69. – P. 113–137. DOI: 10.4000/asp.4843.

[5] Charles, M. Getting the corpus in front of the students: ESP, corpus and student-centred learning // *ASp. La revue du GERAS*. – 2014. – № 66. – P. 9–27. DOI: 10.4000/asp.4527.

[6] Chen, J., Dai, J., Zhu, K., & Xu, L. Effects of extended reality on language learning: A meta-analysis // *Frontiers in Psychology*. - 2022. - Vol. 13. - P. 1016519. DOI: 10.3389/fpsyg.2022.1016519.

[7] Parmaxi, A., & Demetriou, A.A. Augmented reality in language learning: A state-of-the-art review of 2014-2019 // *Journal of Computer Assisted Learning*. – 2020. – Vol. 37, № 4. - P. 861-883. DOI: 10.1111/jcal.12486.

[8] Жубанова Ш.А., Джусубалиева Д.М. Интерактивно-иммерсивный подход в формировании иноязычной профессионально-ориентированной компетенции студентов технического профиля // *Известия КазУМОиМЯ им. Абылай хана. Серия «Педагогические науки»*. – 2025. – № 1(76). - С. 339-358. DOI: 10.48371/PEDS.

[9] Zhou, Q. Principles and Strategies for Improving Intercultural Communication Competence of Senior High School English Students Under the Guidance of Core Competence // *Education Theory: Teaching and Learning*. – 2023. – Vol. 2, № 5. - P. 45-57.

[10] Radianti, J., Majchrzak, T.A., Fromm, J., & Wohlgenannt, I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda // *Computers & Education*. – 2020. – Vol. 147. – P. 103778. DOI: 10.1016/j.compedu.2019.103778.

REFERENCES

[1] Konovalova E.N. Professional English in Use: Architecture [Professional English in Use: Architecture]. Textbook for architecture students. - Moscow: Cambridge University Press / MCBA. - 144 p. [in Rus.]

[2] Bezruchko E.N. English for Architects [English for Architects]. Coursebook for students of architecture. - Moscow: ASV Publishing House. - 176 p. [in Rus.]

[3] Lukina L.V. Anglijskij dlya stroitelnyh universitetov [English for Construction Universities]. ESP textbook for construction and architecture majors. - Moscow: MGSU. - 200 p. [in Rus.]

[4] Boulton A. (2016) Integrating corpus tools and techniques in ESP courses. *ASp. La revue du GERAS*. - № 69. - P. 113–137. DOI: 10.4000/asp.4843.

[5] Charles M. (2014) Getting the corpus in front of the students: ESP, corpus and student-centred learning. *ASp. La revue du GERAS*. - № 66. - P. 9–27. DOI: 10.4000/asp.4527. [in Eng.]

[6] Chen J., Dai J., Zhu K., Xu L. (2022) Effects of extended reality on

language learning: A meta-analysis. *Frontiers in Psychology*. - Vol. 13. - P. 1016519. DOI: 10.3389/fpsyg.2022.1016519.

[7] Parmaxi A., Demetriou A.A. (2020) Augmented reality in language learning: A state-of-the-art review of 2014–2019. *Journal of Computer Assisted Learning*. - Vol. 37, № 4. - P. 861–883. DOI: 10.1111/jcal.12486.

[8] Zhubanova Sh.A., Dzhusubalieva D.M. (2025) Interaktivno-immersivnyy podkhod v formirovaniy inoyazychnoy professionalno-oriyentirovannoy kompetentsii studentov tekhnicheskogo profilya [Interactive-immersive approach in the formation of foreign language professionally oriented competence of engineering students]. *Izvestiya KazUMOiMYa im. Abylai khana. Seriya «Pedagogicheskiye nauki» - Bulletin of Ablai Khan KazUIR&WL. Series «Pedagogical Sciences»*. - 2025. - № 1(76). - P. 339–358. DOI: 10.48371/PEDS. [in Rus.]

[9] Zhou Q. (2023) Principles and Strategies for Improving Intercultural Communication Competence of Senior High School English Students Under the Guidance of Core Competence. *Education Theory: Teaching and Learning*. - Vol. 2, № 5. - P. 45–57.

[10] Radianti J., Majchrzak T.A., Fromm J., Wohlgenannt I. (2020) A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*. - Vol. 147. - P. 103778. DOI: 10.1016/j.compedu.2019.103778.

КӘСІБИ БАҒЫТТАЛҒАН ШЕТ ТІЛІН ОҚЫТУДАҒЫ СӘУЛЕТ ТЕРМИНОЛОГИЯСЫН КОМПЬЮТЕРЛІК ТАЛДАУДЫ ПАЙДАЛАНУДЫҢ ӘДІСТЕМЕЛІК АСПЕКТІЛЕРІ

*Смагулова Ж.А.¹, Жылтырова Ж.Т.², Калелова И.М.³

^{*1,2}Абылай хан атындағы Халықаралық қатынастар және әлем тілдері
университеті, Алматы, Қазақстан

³Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті,
Өскемен, Қазақстан

Аңдатпа. Бұл мақалада болашақ сәулетшілерге арналған арнайы мақсаттағы ағылшын тілі (ESP) курстарындағы сәулеттік терминологияның үш деңгейде - бастапқы (A1), бастапқы-орта (A2) және орта (B1) - таралуы зерттеледі. Зерттеудің эмпирикалық негізін кәсіби бағытталған үш оқулықтан - *Professional English in Use: Architecture* (Коновалова Е.Н.), *English for Architects* (Безручко Е.Н.) және *English for Construction Universities* (Лукина Л.В.) - алынған мәтіндерден құрастырылған арнайы корпус құрайды. Корпус мәтіндері цифрланып, регистрді нормализациялау, стоп-сөздерді алып тастау және токенизациялау арқылы алдын ала өңделді, одан кейін Python, NLTK және gensim кітапханаларын пайдалана отырып жиілік

және тақырыптық модельдеу талдауына ұшырады. Жиілік таралымдары мен визуализациялар - баған диаграммалары, дөңгелек диаграммалар және ruLDAvis тақырыптық карталары - сәулеттік лексиканың жоғары жиілікті тығыз өзегін (architecture, building, construction, materials, design), үш тақырыптық кластерді (құрылымдық элементтер, материалдар, жобалау принциптері) және корпусстың тұжырымдамалық ядросын білдіретін бір басым LDA тақырыбын анықтауға мүмкіндік берді. Жиілік деректері, тақырыптық мүшелік және контекстік қолданыс үлгілері негізінде A1–B1 континуумы үшін терминологияны деңгейлік бөлудің дифференцирленген моделі ұсынылды. Бұл модель VR негізіндегі оқу тапсырмаларын жобалауда қолданылады: A1 деңгейінде виртуалды ортадағы базалық кеңістіктік элементтерді атаудан бастап, B1 деңгейінде форма, стиль және тұрақтылық тұжырымдамаларын біріктіретін күрделі жоба тапсырмаларына дейін. Зерттеу нәтижелері корпусстық-лингвистикалық әдістер мен иммерсивтік VR-технологияларын ұштастырудың сәулеттік лексиканы меңгерудің эмпирикалық тұрғыдан дәйектілеу жолын қамтамасыз ететінін, лексикалық таңдауды аутенттік жиілік профильдері мен академиялық мәтіндердің тақырыптық құрылымымен сәйкестендіретінін және осылайша ESP оқытудың жалпы тиімділігін арттыратынын көрсетеді.

Тірек сөздер: компьютерлік корпусқа негізделген талдау, сәулет терминологиясы, арнайы мақсаттағы ағылшын тілі (ESP), жиілік талдауы, LDA тақырыптық модельдеуі, VR-технологиялар негізіндегі тіл үйрету, құрылымдық элементтер, дизайн қағидаттары

МЕТОДИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ КОМПЬЮТЕРНОГО АНАЛИЗА АРХИТЕКТУРНОЙ ТЕРМИНОЛОГИИ В ОБУЧЕНИИ ПРОФЕССИОНАЛЬНО ОРИЕНТИРОВАННОМУ ИНОСТРАННОМУ ЯЗЫКУ

*Смагулова Ж.А.¹, Жылтырова Ж.Т.², Калелова И.М.³

^{*1,2}Казахский университет международных отношений и мировых языков
имени Абылай хана, Алматы, Казахстан

³Восточно-Казахстанский технический университет имени Д.Серикбаева,
Усть-Каменогорск, Казахстан

Аннотация. В настоящей статье исследуется распределение архитектурной терминологии в курсах английского языка для специальных целей (ESP), разработанных для будущих архитекторов, на трёх уровнях владения языком: элементарном (A1), предпороговом (A2) и пороговом (B1). Эмпирическую основу исследования составляет специально созданный корпус, включающий тексты из трёх профессионально ориентированных учебников: *Professional English in Use: Architecture* (Коновалова Е.Н.),

English for Architects (Безручко Е.Н.) и *English for Construction Universities* (Лукина Л.В.). Тексты корпуса были оцифрованы и предварительно обработаны посредством нормализации регистра, удаления стоп-слов и токенизации, после чего подвергнуты частотному анализу и тематическому моделированию с использованием библиотек Python, NLTK и gensim. Частотные распределения и визуализации - столбчатые и круговые диаграммы, тематические карты pyLDAvis - позволили выявить компактное высокочастотное ядро архитектурной лексики (architecture, building, construction, materials, design), три тематических кластера (конструктивные элементы, материалы, принципы проектирования) и одну доминирующую LDA-тему, представляющую концептуальное ядро корпуса. На основе частотных данных, тематической принадлежности и контекстуальных паттернов употребления предложена дифференцированная по уровням модель распределения терминологии для континуума А1–В1. Данная модель применяется при проектировании VR-заданий: от называния и описания базовых пространственных элементов в виртуальной среде на уровне А1 до сложных проектных заданий, интегрирующих понятия формы, стиля и устойчивого развития, на уровне В1. Полученные результаты свидетельствуют о том, что сочетание корпусно-лингвистических методов и иммерсивных VR-технологий обеспечивает более эмпирически обоснованную траекторию освоения архитектурной лексики, согласуя лексический отбор с аутентичными частотными профилями и тематической структурой академических текстов, тем самым повышая общую эффективность обучения ESP.

Ключевые слова: компьютерный корпусный анализ, архитектурная терминология, английский язык для специальных целей (ESP), частотный анализ, тематическое моделирование LDA, обучение иностранному языку с использованием VR, структурные элементы, принципы дизайна

Received / Мақала түсті / Статья поступила: 29.12.2025
Accepted / Жариялауға қабылданды / Принята к публикации: 26.06.2026

Information about the authors:

Smagulova Zhadyra Alimbekovna - PhD student, Kazakh Ablai Khan University of International Relations and World Languages, Almaty, Kazakhstan, e-mail: jadirasmagulova1989@gmail.com

Zhylytyrova Zhanar Tekezbaykyzy – Vice-Rector for International Relations, Associate Professor, Kazakh Ablai Khan University of International Relations and World Languages, Almaty, Kazakhstan, e-mail: zhylytyrova.zh@ablaikhan.kz

Kalelova Indira Maulkanovna - Senior teacher, Master of the Pedagogical Sciences, Language Competences Centre of D. Serikbayev EKTU, Oskemen, Kazakhstan, e-mail: indira.kalelova@mail.ru

Авторлар туралы мәлімет:

Смағұлова Жадыра Алимбековна - PhD докторант, Абылай хан атындағы Қазақ халықаралық қатынастар және әлем тілдері университеті, Алматы, Қазақстан, e-mail: jadirasmagulova1989@gmail.com

Жылтырова Жанар Текесбайқызы – халықаралық байланыстар жөніндегі проректор, қауымдастырылған профессор, Абылай хан атындағы Қазақ халықаралық қатынастар және әлем тілдері университеті, Алматы, Қазақстан, e-mail:zhyltyrova.zh@ablaikhan.kz

Калелова Индира Маулкановна – Д.Серікбаев атындағы ШҚТУ Тілдер құзыреттілігі орталығының аға оқытушысы, педагогика ғылымдарының магистрі, e-mail:indira.kalelova@mail.ru

Информация об авторах:

Смағұлова Жадыра Алимбековна - PhD докторант, Казахский университет международных отношений и мировых языков имени Абылай хана, Алматы, Казахстан, e-mail: jadirasmagulova1989@gmail.com

Жылтырова Жанар Текесбайқызы – проректор по международным связям, ассоциированный профессор, Казахский университет международных отношений и мировых языков имени Абылай хана, Алматы, Казахстан, e-mail: zhyltyrova.zh@ablaikhan.kz

Калелова Индира Маулкановна - старший преподаватель, магистр педагогических наук Центра языковых компетенций ВКТУ им.Д.Серикбаева, Усть-Каменогорск, Казахстан, e-mail: indira.kalelova@mail.ru