



REVIEW ARTICLE

Section: *Literature, Linguistics & Criticism*

Corpus-based Arabic language pedagogy for Turkish learners: A bilingual Arabic–Turkish model for data-driven learning

Muhammet Abazoglu¹, Saif Al Mahrouqi^{2*}, Hanene Lahiani³, Saeed Nasser Alrayssi³, Mohammed Abou Adel², Ahmad M. Al Mahamed⁴

¹College of Humanities and Social Sciences, Department of Arabic Language and Literature, Kilis 7 Aralık University, Kilis, Türkiye

²College of Humanities and Social Sciences, Department of Arabic Language and Literature, UAE University, Al Ain, UAE

³College of Education, English Language and Translation Program, Al Ain University, Abu Dhabi, UAE

⁴College of Arts, Education and Social Sciences, Department of Communication, Abu Dhabi University, Abu Dhabi, UAE

*Correspondence: *Saif Al Mahrouqi*

ABSTRACT

Linguistics has experienced a major transition over the past few decades, shifting from predominantly structural approaches to data-driven approaches based on authentic language use. This development has strengthened the role of computational linguistics within traditional linguistic disciplines. Despite the growing global interest in teaching Arabic as a Foreign Language (AFL), instructional materials continue to rely heavily on artificial language samples that do not accurately reflect real communicative practices. This study investigates the pedagogical benefits of corpus-based text sets for Turkish learners of Arabic. The research adopts a mixed-methods approach and employs the following procedures: first, it identifies and analyzes the pedagogical benefits of corpus-based instruction; second, it compares learning outcomes using a bilingual Arabic–Turkish corpus containing approximately 100,000 words drawn from literary texts representing classical, modern, and contemporary periods. AntConc software, which enables comprehensive corpus-analysis through concordances and collocation tools, is employed to examine frequency distributions, collocational patterns, n-grams, and contextual concordances, thereby providing a systematic and replicable corpus-informed analytical framework. The findings demonstrate, through illustrative examples, how discourse is semantically and pragmatically structured, as evidenced by common collocations associated with the word “president” in political discourse and the diverse cultural and social contexts surrounding the word “Arabic.” The results indicate that corpus-based text sets support vocabulary development, enhance grammatical competence, and strengthen communicative abilities by exposing learners to authentic linguistic input rather than decontextualized language examples. Furthermore, the study contributes to the integration of corpus linguistics into AFL teaching by proposing a data-driven pedagogical framework and emphasizing the need for corpus-informed teacher training. It also highlights the value of interdisciplinary collaboration between computational linguistics and language education, positioning corpus-based instruction as a scalable, empirically grounded, and pedagogically robust approach that responds to current developments in applied linguistics.

KEYWORDS: Corpus linguistics, Arabic pedagogy, Turkish learners, computational linguistics, data-driven learning

Research Journal in Advanced Humanities

Volume 7, Issue 4, 2026

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: 30 June 2026

Accepted: 23 August 2026

Published: 15 September 2026

HOW TO CITE

Abazoglu, M., Al Mahrouqi, S., Lahiani, H., Alrayssi, S. N., Adel, M. A., & Al Mahamed, A. M. (2026). Corpus-based Arabic language pedagogy for Turkish learners: A bilingual Arabic–Turkish model for data-driven learning. *Research Journal in Advanced Humanities*, 7(4). <https://doi.org/10.58256/hv7d8031>



Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

© 2026 The Author(s). This is an open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Teaching Arabic as a foreign language has undergone significant developments in recent years. Arabic is widely recognized as a language of considerable, scientific, professional, and cultural importance (Abouelnour et al., 2025; Adel et al., 2024b). This has prompted researchers to investigate new methods and technologies for teaching this language (Alhourani et al., 2026; Qiqieh et al., 2025). Accordingly, computational linguistics serves as a bridge between linguistics and computer science, offering advanced tools and approaches to language acquisition (Shaaban et al., 2024; Abouelnour et al., 2024b).

The introduction of language corpora has provided researchers and instructors with large and comprehensive collections of authentic texts that reflect real-world language use. These texts differ from traditional, artificially constructed examples, because they provide context-rich linguistic data that enable learners and researchers to observe authentic usage patterns across a wide range of communicative settings. This learning method enables learners to recognize frequently used vocabulary, recurrent grammatical structures, and collocational patterns while simultaneously understanding the pragmatic and cultural dimensions of language use (Abouelnour et al., 2024b). As a result, learners acquire deeper linguistic knowledge and better communication skills for real-life situations (Adel et al., 2024a).

1.1. Theoretical Foundations of Corpora

In Arabic, a corpus (مجموعة) or mudawwana derives from the root dawwana, which means to collect, organize, and document information. Its Latin equivalent *corpus* literally means ‘body’ and is widely used in European languages to refer to large collections of texts. In modern linguistics, this term refers to a systematically compiled collection of written or spoken texts that has been digitally processed and stored for a linguistic analysis (Al-Dekrori, 2018; Al-Zuhairi, 2003).

Al-Faifi (2022) demonstrates that morpho-syntactic annotation of Arabic corpora benefits both scholars and educators, enabling the latter to identify regularities of lexis and semantics that are related to language practice, thereby facilitating the creation of more accurate educational programs. The importance of the corpora in terms of lexicography and terminology normalization is further highlighted by Dalool (2014).

Although corpus-based analysis methodologies have been fully embraced for decades in Western language teaching (Boulton, 2017), their application to Arabic language teaching remains comparatively limited, largely because of the scarcity of large, annotated Arabic corpora (Zeroual et al., 2018). This shortage restricts the broader application of computational tools in teaching Arabic, thus opening the door for further research to address the gap.

1.2. Corpus Linguistics Within Computational Linguistics

The convergence of knowledge of linguistics and advances in technologies gave rise to computational linguistics as a vital field of applied linguistics in the present digital era. With better computing power and an evolution in AI techniques, computational linguistics becomes a field of investigation that is witnessing a steep rise in its development and is crucially involved in developing systems for human-machine linguistic interactions (Algobaei & Alzain, 2026; Mohamed et al., 2025; El-Erian et al., 2024; Abdullah & Alzain, 2026). This field constitutes the foundation of defining areas like natural language processing, machine learning, machine translation, discourse analysis, and big data handling according to Al-Anati (2005).

Text linguistics is an immediate application of computational linguistics, as experiments have proven its effectiveness when used in teaching Arabic to foreigners. The usage of different corpora makes students aware of the patterns in the language and helps them learn grammatical structures through noticing and discovery, instead of dealing with mere abstract rules (Abd al-Ghani, 2024). Latest research in machine translation demonstrates that handling a complicated language such as Arabic demands several strategies because of the differences in sentence structure patterns, the multiplicity of meanings, and the complexity of morphological structure. Therefore, it requires accurate morphological analysis as well as training rules developed specifically for the target language (Norqobilov, 2024).

In this context, Brezina (2018) sees text linguistics as a descriptive approach, which is based on actual linguistic evidence. This method has been instrumental in developing new teaching practices, the most significant of which are vocabulary selection through frequency, syntactic awareness enhancement, and digital material

creation that resembles the natural use of language (Al-Fifi, 2023; McEnery & Hardie, 2012). Moreover, the impact of text linguistics has widened to include phonetics, morphology, lexicography, semantics, pragmatics, and discourse analysis (Al-Shahri, 2022). In this regard, corpus linguistics, whether directed to language itself or used as a tool in the study of language, is not merely a linguistic theory but rather a new method in linguistic research that differs from traditional approaches (Abazoglu & Alhourani, 2025).

1.3. Pedagogical Relevance and the Arabic–Turkish Context

Using corpus linguistics in teaching Arabic as a foreign language, specifically within the Turkish-medium educational context (where Arabic is gaining popularity), is still at a relatively nascent stage of progression. At present, traditional syllabi still depend on fabricated sentences or single instances that are not based on authentic language use. As a result, they not only fail to equip students with pragmatic skills but also rarely enable them to comprehend language in context. Conversely, data-driven learning (DDL) empowers students to explore language through authentic models, whereby they personally extract grammar and standards from real usage. This method allows Turkish learners to overcome formidable challenges that emerge from deep-rooted differences between Arabic and Turkish, such as root-based morphology, sentence structure, and semantic context. Additionally, the availability and use of bilingual (Arabic–Turkish) corpora enable learners to draw parallels between these languages, improve linguistic knowledge, and acquire and internalize frequent linguistic patterns. Thus, a curriculum designed through utilizing corpora is a revolutionary and efficient way of teaching Arabic at universities and language-learning centers across Turkey.

2. Method

2.1. Research Design

The study employs a descriptive-analytical methodology that combines theoretical and practical components. Theoretically, the paper explores linguistic corpora's structure, types, and functions in the learning process. In addition to defining these elements as a part of computational linguistics, the study explains their role in teaching Arabic to non-native speakers (Al-Dekrouri, 2018; Al-Zuhairi, 2003). The foundations that transform corpora into a significant source for researchers and educators are also a major point of discussion in this study, including the importance of the scientific value of morpho-syntactically annotated Arabic corpora, as documented by Al-Faifi (2022).

The research focuses on computational linguistics, which is a middle path between traditional linguistic analysis and the latest digital technologies (Adel, 2022). Since corpus-based tools are considered among the most significant applications in the field, this study is based on the assumption that these tools enable the investigation of authentic language use through comprehensive and reliable data (Brezina, 2018; McEnery & Hardie, 2012). Consequently, a bilingual Arabic–Turkish corpus is analyzed to examine grammatical and semantic patterns as well as lexical frequencies to derive cohesive relationships in the texts and infer educational implications for Arabic teaching (Khater, 2023).

2.2. Corpus Construction and Data Processing

The study corpus was designed to comprise approximately 100,000 words, and was intended to serve as an authentic linguistic resource for language analysis and educational applications. The corpus consisted of two parts:

(a). The Arabic Section

This section includes examples that reflected specific stages in the development of Arabic as well as various aspects of the spoken language:

- Poetry from the pre-Islamic era
- Prose from the classical age
- Literary works from the Umayyad and Abbasid periods
- Samples of modern and contemporary literature
- Journalistic writing and political as well as cultural speeches

Texts selected as representative samples were also characterized by their historical importance and usefulness in studying authentic usage of Arabic. Texts with a well-established authorship and obvious linguistic patterns were prioritized, while repetitive and incomplete ones were not included in order to maintain consistency in the corpus.

(b). The Turkish Section

This section comprises Arabic texts along with their Turkish counterparts in a parallel form as well as Turkish texts which share the rhetorical similarity, thus giving the learners an opportunity to see how the two languages interact with each other in the process.

The subsequent section includes preparation and processing of data, comprising the following operations:

- Elimination of texts containing non-linguistic elements such as subtitles, pagination, etc.
- Standardization of texts in spelling, punctuation, and character encoding so that the texts were consistent and logically coherent
- Metadata information was recorded for each text, including the author's or translator's identity, publication date, source, and literary genre

Additionally, the following corpus design principles were considered:

- Representation—inclusion of various text types
- Balance—the texts were distributed without the dominance of any single type
- Authenticity—the material used was unaltered and thus reflected actual language usage

This annotation made the corpus more transparent, which would allow it to be reused in subsequent research. Furthermore, it facilitated the analysis at different linguistic levels.

Figure 1 summarizes the corpus-informed workflow adopted in this study, showing how authentic Arabic-Turkish textual data are selected, prepared, annotated, analyzed, and transformed into pedagogical applications for Turkish learners of Arabic.



Figure 1 Workflow of the Arabic-Turkish Corpus-Based Pedagogical Model

2.3. Analytical Tools and Procedures

The corpus data were analyzed with AntConc 4.0, which is one of the most widely used open-source software packages for corpus and text analysis. The four primary instruments used were as follows:

- Word list; finding lexical trends and extracting word frequencies
- Collocation: comprehending how words relate to one another verbally in various settings

- Concordance; analyzing the terminology's contextual use and semantic variety
- N-gram models; demonstrating grammatical rules and recurrent syntactic patterns

These tools were selected for their performance capabilities, which have proven to be compelling in quantitative and qualitative linguistic research, particularly in text-based applications (Baker, 2014; McEnery & Hardie, 2012).

2.4. Ethical Considerations

All the texts used in this study are openly accessible and were collected according to the established scientific and ethical standards without any direct contact with people. The figures (2–4), in the subsequent section, illustrate the results and show how frequently N-gram models occur and correlate with each other.

3. Results

This section presents the outcomes that were derived in the research process while analyzing the Arabic–Turkish corpora based on the AntConc software. Data analysis involved distinguishing the lexical, grammatical and semantic features of the corpus and was carried out based on four principal analytical methods: frequency analysis, collocation, contextual index analysis, n-gram analysis (Abou Adel et al., 2025).

3.1. Frequency Results

The frequency list created by AntConc's word list tool for both Turkish and Arabic texts proved that the words with grammatical functions underwent the most frequent usage in the Arabic corpus. This means that the construction of Arabic sentences necessitates making use of function words.

Frequent auxiliary verbs such as etmek-olmak, etmiştir, etc., can be observed in the Turkish part of the syntax, along with relevant prepositions such as -den -dan. Hence, it can be concluded that texts used in the Turkish and Greek sections of the corpus reflect similar themes in politics, religion, and society.

3.2. Collocation Analysis

The collocation tool was applied to determine which words statistically accompany the word president (الرئيس) in the corpus. The findings highlight several frequently occurring collocations, such as republic (الجمهورية), government (الحكومة), and ministers (الوزراء), all of which denote the institutional and political character of the contemporary texts that typically concern administration and governance.



Collocate	Rank	FreqL	FreqR	Range	Likelihood	Effect
1 الجمهورية	1	3	0	3	22.962	6.862
2 الحكومة	1	3	0	3	22.962	6.862
3 الوزراء	1	3	0	3	22.962	6.862
4 عام	4	2	0	2	14.772	6.072

Figure 2: Examples of Sentences That Contain the Word President (الرئيس)

This pattern mirrors the grammatical framework of formal Arabic speech and is a major resource for training students to express their ideas in politics and the media through authentic examples.

3.3. Concordance Analysis

Analysis of the term “Arabic” (العربية) in its various meanings in related segments of the texts was conducted. It shows that the term has various meanings within cultural, social, and political spheres as exemplified by phrases

teaching vocabulary and developing contextual semantics.

These findings demonstrate that corpus-based approaches depict usage patterns in natural language, which may then be used to develop instructional resources that are rooted in a real-world context and actual usage.

4. Discussion

4.1. Integrating Language Corpora into Foreign Language Teaching

The findings of this study indicate that language corpora should not only be considered as mere data repositories but as educational instruments that enable students to transition from intuitive to evidence-based learning styles, which aligns with the viewpoints of various researchers (Oktavianti, 2021; Sridhar, 1993). Using authentic texts in teaching AFL logically leads learners to acquire not only linguistic but also contextual and communicative competence as they become familiar with the language in real-life situations rather than only hearing artificially created examples that do not reflect language realities (O'Keeffe & McCarthy, 2022).

The results demonstrate limited corpora usage in Arabic language teaching, the main reason for which is the lack of teachers' digital skills and the absence of sufficient specialized training (Oktavianti, 2021). The study utilized a bilingual Arabic–Turkish corpus, addressing these problems by offering an easily accessible bilingual language resource that assists teachers in language instruction while prioritizing learners, as activities based on real language use can be more easily designed to integrate student practice with DDL (Szudarski, 2022).

Hence, this research shows that incorporating any kind of text into teaching depends upon institutional support, systematic training programs, and local corpora that consider the learners' linguistic–educational backgrounds. The Arabic–Turkish corpus exemplifies a tool that not only allows teachers to understand text linguistics as a theoretical framework but also enables them to practically achieve classroom implementation of text linguistics.

4.2. The Influence of Language Texts on Building the Four Language Skills

The investigation reveals that the Arabic–Turkish corpus provided valuable material for developing the four fundamental language skills: listening, speaking, reading, and writing. Speaking materials familiarized learners in a natural way with the most common speech patterns so that they could effectively improve their listening and oral interaction skills outside the scope of traditional textbook models (Djalilova, 2024). Furthermore, the range of daily expressions and grammatical structures raised the learners' understanding of pragmatic and stylistic variations, a point that aligns with the findings of multimedia research, which argue that the simultaneous use of textual and audio input is the most effective method to advance skills in the interactive mode (Titova & Ignatova, 2024).

As for reading skills, the results of repetition and juxtaposition activities reveal that corpus-based input is instrumental for educators to discover core vocabulary and frequently occurring grammatical structures, which enables them to craft linguistically accurate reading materials (Tracy-Ventura & Braidi, 2023). A corpus additionally allows learners to follow the spread of terms in different texts, such as human rights (حقوق الإنسان) or sustainable development (التنمية المستدامة) which connects reading to real cultural and social contexts.

As for writing, the most difficult productive skill, the findings indicate that corpora are a powerful means for demonstrating real-world writing structures and vocabularies. Unlike dictionaries, corpora offer learners the correct instances of sentence structures, thereby requiring them to improve their stylistic and grammatical accuracy (Hassan, 2023). The statistics reveal that students can identify regular patterns, such as making a decision (اتخاذ قرار) or making a suggestion (التقديم المقترح) and can subsequently use them naturally in their compositions. This aligns with research findings that point to DDL's effectiveness in enhancing fluency and lexical accuracy (Szudarski, 2022; Keck, 2012).

Overall, these outcomes highlight that using corpora in instruction helps with guided exploratory learning. At first, teachers guide students through the corpus tools and, subsequently, students start conducting the analysis on their own, and then they draw the relevant conclusions independently. This does not really contradict constructivist learning principles, because they claim that the deepest and longest-lasting learning comes from authentic input plus inductive reasoning (Hizbullah et al., 2020).

Overall, these outcomes emphasize that employing corpora in instruction fosters guided exploratory

learning. Initially, teachers lead students through the corpus tools, and subsequently, students become capable of performing the analysis and drawing the conclusions independently. This does not contradict constructivist learning principles, which hold that the deepest and longest enduring learning is the result of authentic input and inductive reasoning (Hizbullah et al., 2020).

4.3. Construction of a Corpus and Criteria

This research highlights the importance of representativeness, balance, and accuracy in the encoding of a language corpus in determining its quality. This corpus was constructed according to the principles of text linguistics and in accordance with the guidelines of Sinclair (1991) and Al-Juraishi (2011). The corpus consists of works of fiction, newspaper, magazine articles, and other forms of modern writing. Therefore, it provides a wide range of different styles and patterns while maintaining a high degree of internal consistency to facilitate comparability.

The corpus has been annotated for its language features (morphology, syntax, and semantics), and this has made up the analytic resources of the corpus with the metadata attached to it (author, topic, and time), allowing for cross-genre and/or cross-period comparisons. The large difference between classical and modern texts in the Arabic corpus provides examples of the richness of language and demonstrates both research and educational use of the language (Al-Faifi, 2022).

According to the research outcomes, one very important aspect of ethical and practical concerns is that publishers must respect all copyright laws and provide appropriate representations of various cultures through textbook selections used to form the corpus. The ethics and practical issues support the reliability of the corpus as a safe educational tool (Al-Dekrori, 2018). The recommendations made by the authors, when adhered to, will form the basis for the development of future educational corpora that can include language learner data or dialects.

4.4. Practical Instruments and Bilingual Implementation

AntConc has three basic functions (word frequency lists, collocates, and concordances) that provide a practical application of the program in the classroom. Students studied Arabic–Turkish data by looking at frequency counts, collocation patterns, and contextual usage.

One study was conducted to figure out the collocations of the term president (الرئيس), and it showed a rather clear relationship with wording connected to the institution, like republic (الجمهورية), ministry (الوزارة), and government (الحكومة), in a way that kind of mirrored the current political discourse. For instance, the word Arabic was contextually matched to portray its far-reaching effects on politics, culture, and identity, such as the Arab family (الأسرة العربية) and the Arab strategy for the elderly (الاستراتيجية العربية لكبار السن). These outcomes provide a foundation for the progress of data-driven speaking exercises that focus on grammatical patterns, thematic vocabulary, and structures in real-world situations.

These results support the view that corpora facilitate linguistic awareness and learner autonomy (Candarlı, 2022; Roberto, 2023). The Arabic–Turkish corpus is an innovative factor that opens the way for learners to directly compare the two languages, develop translation skills, and understand the structural differences. Using a tool such as arabiCorpus also improves morphological awareness, especially regarding roots such as kataba (كتب). Using a tool such as arabiCorpus also improves morphological awareness, especially regarding roots such as kataba (كتب), a root that generates various words related to writing, such as “book,” “writer,” and “office.”

Therefore, corpus tools should not be seen as programs for analysis only but rather as educational media that bridge authentic linguistic data with classroom practice and consequently change Arabic language teaching from a traditional, rote-learning model to a research-based, inductive, and observational approach.

4.5. Research Contribution, Gaps, and Limitations

This research is a distinct significant contribution in that it introduces the first structural model that is specifically tailored for an educational Arabic–Turkish corpus. Whereas scholars generally provide linguistic descriptions of the corpora, this paper integrates the analysis with education using AntConc software to offer a practical framework for transforming authentic data into classroom tasks.

The findings reveal the following potentials of corpus-based pedagogy:

Developing expressive competence and fluency (Sysoyev & Klochikhin, 2022)

Supporting vocabulary acquisition through repetitive tasks (Szudarski, 2022)

Improving grammatical accuracy and morphological awareness (Hassan, 2023)

Increasing cultural awareness via authentic expressions and proverbs (Al-Ghamdi & Atwell, 2019)

Moreover, the research reveals that learning through corpora is not exclusively for advanced-level students; intermediate learners also find advantages if they are gradually introduced to the corpora. Therefore, it is possible to extend the range of applicability to different levels. Nevertheless, the study's scope is restricted, and many classroom experiments are needed to determine how its effects change over time. Furthermore, the corpus needs new abilities like continuous integration into NLP software or adaptive learning lines, which could be the subject of further research. Last but not least, the corpus is not exhaustive in terms of Arabic dialects.

Finally, it can be stated that the data show that corpus-based language learning perfectly exemplifies the interconnection between linguistic theory and teaching practice and offers an evidence-based model of Arabic teaching that could open up new perspectives for further applied research.

5. Conclusion

The research demonstrates that students can improve their communication skills. In addition, one can improve grammatical knowledge and vocabulary through bilingual texts and AntConc software, which provide access to real-life language data. It also demonstrates that the use of real-life data results enhances the understanding of common lexical and pragmatic structures, hence motivating the development of autonomous language skills via scientific induction. Additionally, the results obtained from the study show that the structural aspects of the corpus like representation, balance, and correctness are vital because they provide the foundation upon which the corpus can be utilized for instruction. The growing demand for materials that can address the linguistic needs of the Turkish-speaking learners, who were previously ignored, can be deemed as the reason for the suggested Arabic-Turkish bilingual approach, as a way to start the creation of educational materials tailored for the benefit of this particular community. Nevertheless, there are certain weaknesses in the study apart from its benefits. First of all, the effect of the corpus could not be monitored for a prolonged period in more than one course. Furthermore, the application covered a limited geographical area and period of time. In future studies, the need for more prolonged application periods, the development of the corpus taking into account different Arabic dialects and texts in audio and video formats, and the use of AI models should be considered. Finally, the study shows that the use of text linguistics in conjunction with computational techniques can be a successful educational approach to teach the Arabic language. This approach is likely to lead to the creation of better curricula for multilingual students in the future.

Disclosure of AI Use: The authors used OpenAI's ChatGPT (GPT-5.6) to refine language and generate the visual abstracts (Figures), followed by careful human review and revision.

References

- Abazoglu, M., & Alhourani, M. I. (2025). The use of language corpora in teaching Arabic to Turkish speakers within the framework of computational linguistics. *Social Sciences & Humanities Open*, 12. <https://doi.org/10.1016/j.ssaho.2025.101947>
- Abd al-Ghani, A. A. M. (2024). Dawr lisānīyāt al-mudawwanāt fī ta'lim al-lughah al-'Arabiyyah ka-lughah thānīyah [The role of corpus linguistics in teaching Arabic as a second language]. *Majallat Kulliyat al-Lughah al-'Arabiyyah bi-Itay al-Barud*, 37(2), 1085–1134. <https://doi.org/10.21608/jlt.2024.345065>
- Abdullah, H. M. A., & Alzain, E. (2026). AI-driven predictive analytics for early identification of at-risk students academically in Saudi universities. *International Journal of Intelligent Engineering and Systems*, 19(6), 75–89. <https://doi.org/10.22266/ijies2026.0630.05>
- Abdulsaboor, A. (n.d.). Data-driven approach for teaching Arabic as a foreign language. *Journal of College of Education for Women*, 31(1), 1–17. <https://doi.org/10.36231/coedw.v31i1.1325>
- Abouelnour, M. et al. (2025). Grammatical performance and its implications in Farouk Shousha's program Our Beautiful Language. *Forum for Linguistic Studies*, 7(5), 670–681. <https://doi.org/10.30564/fls.v7i5.7791>
- Abouelnour, M. et al. (2024a). Interactive communicative discourse in the Holy Qur'an: A deliberative approach. *Edelweiss Applied Science and Technology*, 8(5), 1819–1830. <https://doi.org/10.55214/25768484.v8i5.1915>
- Abouelnour, M. et al. (2024b). The narrative structure in Ibrahim Nasrallah's novel The Second War of the Dog: An analytical study. *Edelweiss Applied Science and Technology*, 8(6), 5481–5489. <https://doi.org/10.55214/25768484.v8i6.3212>
- Abou Adel, M. A., Elhalafawy, A. I., Khodier, G. A., Alrefai, A. A., & Khalifa, S. H. (2025). Unsettling the norm: A posthumanist reading of sameness and differences in Kathryn Erskine's *Mockingbird*. *Theory and Practice in Language Studies*, 15(11), 3511–3520. <https://doi.org/10.17507/tpls.1511.03>
- Adel, M. A. (2022). Towards an advanced interactive e-learning for the language. *Arab World English Journal (AWEJ) Special Issue on CALL* (8), 330–340. <https://dx.doi.org/10.24093/awej/call8.22>
- Adel, M. A., Mohamed, A. M., & Altwaiji, M. (2024a). The representation of women with disabilities in Shahira Al-Hasan's *Cochlea's Secrets*. *Journal of International Women's Studies*, 26(6), Article 8.
- Adel, M. A., Boudjadi, K., Abouelnour, M. M., & Alhourani, M. I. (2024b). The contribution of smartphone apps to develop teaching the Arabic language 'Arabic Is My Language's App' as a sample. *Forum for Linguistic Studies*, 6(6), 1175–1190. <https://doi.org/10.30564/fls.v6i6.7408>
- Ali, N., Hasanah, M., & Prasetyo, A. (2020). The integration of Qur'an and linguistic education based on ontology of Qur'anic concept in Quranic Arabic corpus. *Journal of Arabic Learning*, 3(2), 146–158. <https://doi.org/10.18860/IJAZARABI.V3I2.9769>
- Algobaei, F., & Alzain, E. (2026). Prompt engineering for non-native English learners: A generative AI approach to personalised language feedback. *Social Sciences & Humanities Open*, 13. <https://doi.org/10.1016/j.ssaho.2025.102341>
- Al-Anati, W. (2005). Arabic computational linguistics: Concepts, applications, and feasibility. *Zarqa Journal for Research and Studies*, 7(2), 61–81. <https://search.emarefa.net/detail/BIM-28844>
- Al-Dekrori, A. (2018). Linguistic corpora and their role in processing Arabic texts. *King Abdullah bin Abdulaziz International Center for the Arabic Language*. https://www.researchgate.net/publication/399123477_Corpus_Linguistics_and_its_role_in_Processing_Arabic_Texts_almdwnat_allghwyt_wdwrha_fy_maljt_alnsws_alrbyt
- Al-Faifi, A. (2023). Linguistic corpus tagging: Concept and fields. *Mu'tah Journal of Research and Studies*, 38(1), 275–316. <https://doi.org/10.35682/mjhsc.v38i1.603>
- Al-Ghamdi, A., & Atwell, E. (2019). Constructing a corpus-informed list of Arabic formulaic sequences (ArFSs) for language pedagogy and technology. *International Journal of Corpus Linguistics*, 24(2), 202–228. <https://doi.org/10.1075/ijcl.16088.alg>
- Alhourani, M. et al. (2026). The impact of xenophobia on foreign language learning anxiety among Turkish university students: The case of Arabic. *Research Journal in Advanced Humanities*, 7(1). <https://doi.org/10.58256/xgv2c354>
- Al-Juraishi, A. (2011). A proposed concept for preparing the necessary corpus linguistics for learners of Arabic

- as a second language in light of the criteria for their construction. *Islamic University Journal of Arabic Language and Literature*, 4(1), 505–552. <https://search.emarefa.net/detail/BIM-1584934>
- Almujaiwel, S. (2018). *Integrating NLP with corpus linguistics and vice versa* [Conference Article]. LOPAL '18: Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications, Rabat, Morocco. <https://doi.org/10.1145/3230905.3230930>
- Al-Rayzah, A., Al-solami, F., & Saleh, M. (2024). AraFast: Developing and evaluating a comprehensive modern standard Arabic corpus for enhanced natural language processing. *Applied Sciences*, 14(12), 5294. <https://doi.org/10.3390/app14125294>
- Al-Shahri, A. (2022). Corpus linguistics: Definition and areas of use. *Arab Journal of Science and Research Publishing*, 8(3). <https://doi.org/10.26389/AJSRP.R030822>
- Al-Zuhairi, M. (2003). *Dictionary of informatics and computational linguistics terms*. Library of Lebanon.
- Arifianto, M. L. (2021). Utilizing the Quranic Arabic corpus as a supplementary teaching and learning material for Arabic syntax: An overview of a web-based Arabic linguistics corpus. *KnE Social Sciences*, 403–412. <https://doi.org/10.18502/kss.v5i3.8563>
- Boulton, A. (2017). Data-driven learning and language pedagogy: Recent developments and future directions. *Language Learning and Technology*, 21(3). https://link.springer.com/rwe/10.1007/978-3-319-02328-1_15-2
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Candarlı, D. (2022). Corpus data and methods. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 52–73). Routledge eBooks. <https://doi.org/10.4324/9781003081890-3>
- Dalool, I. (2014). *Computerized lexicon of trilateral verbs in Arabic* [Master's thesis, Islamic University of Gaza] <https://ebook.univeyes.com/112698>
- Djalilova, A. N. (2024). Use of spoken corpora in language learning and teaching. *Texas Journal of Philology, Culture and History*, 28, 4–7. <https://pdfs.semanticscholar.org/a7d3/bd73e9806bb5bcbecee2f2ccc2349ab9e560>
- El-Erian, M., Imad, D., Sulaiman, S., Qutieshat, E., & Khater, M. (2024). The role of artificial intelligence in enhancing corporate governance and achieving sustainable development. *Access to Justice in Eastern Europe*, 9(1). <https://doi.org/10.33327/AJEE-18-9.1-a000177>
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, 20(3), 465–480. <https://doi.org/10.1558/cj.v20i3.465-480>
- Granger, S. (2008). Learner corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook*. Walter de Gruyter. <https://doi.org/10.1002/9781405198431.wbeal0669>
- Hassan, H. (2023). Applications of linguistic corpora in teaching and learning Arabic as a foreign language. *Publications of the Islamic World Educational, Scientific and Cultural Organization (ICESCO)*.
- Hizbullah, N., Arifa, Z., Suryadarma, Y., Hidayat, F., Muhyiddin, L., & Firmansyah, E. K. (2020). Source-based Arabic language learning: A corpus linguistic approach. *Humanities and Social Sciences*, 8(3), 940–954. <https://doi.org/10.18510/hssr.2020.8398>
- Keck, C. (2012). Corpus linguistics in language teaching. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405198431.wbeal0256>
- Khater, M. (2023). Refugee children's right to education: The education of Syrian refugee children in Jordan—Reality and prospects. *Access to Justice in Eastern Europe*, 6(3), 108–119. <https://doi.org/10.33327/AJEE-18-6.3-a000302>
- Sysoyev, P. V., & Klochikhin, V. V. (2022). Development of students' collocational competence based on corpora. <https://doi.org/10.32744/pse.2022.4.19>
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Mohamed, A. M., Mouas, S., Abdelaty, Y., & Yusuf, A. (2025). AI technologies and multicultural education insights into teachers' attitudes and challenges in higher education. In *Smart business and education in the age of VUCA and BANI* (pp. 153–178). <https://doi.org/10.4018/979-8-3373-7922-7.ch007>
- Norqobilov, S. (2024). Comparing computational linguistics approaches across languages. *Deleted Journal*, 3(12/2). <https://www.elibrary.ru/item.asp?id=73823461>
- O'Keeffe, A., & McCarthy, M. J. (2022). *The Routledge handbook of corpus linguistics*. Taylor & Francis.
- Oktavianti, I. N. (2021). *Corpora: From theoretical linguistics to language teaching*. UAD TEFL International

- Conference, 2, 19. <https://doi.org/10.12928/utic.v2.5731.2019>
- Qiqieh, S., et al. (2025). The impact of WhatsApp on first-year students' transition to university life. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1552278>
- Roberto, T. M. G. (2023). Corpus linguistics in language teaching. *Revista Foco*, 16(7), e2631. <https://doi.org/10.54751/revistafoco.v16n7-089>
- Salabai, S.-I. (2025). *The role of language corpora in foreign language teaching: Trends, debates, and applications*, Proceedings of the ICND conference, (05/30/2025; Kiev, Ukraine), 490–492. <https://doi.org/10.62731/mcnd-30.05.2025.015>
- Saleh, M. I. (2015). Linguistic corpus and how to benefit from them. In *Arabic linguistic corpora: Their structure and methods of benefiting from them*.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Shaaban, T. S., Mohamed, A. M., Nasim, S. M., Janada, R. A., Alfaisal, A., & Mujeeba, S. (2024). Promoting multilingualism and executive functions: A cross-cultural study on language learning support. *XLinguae*, 17(2), 200–215. <https://doi.org/10.18355/XL.2024.17.02.14>
- Sridhar, S. N. (1993). What are applied linguistics. *International Journal of Applied Linguistics*, 3(1), 3–16. <https://doi.org/10.1111/j.1473-4192.1993.tb00040.x>
- Szudarski, P. (2022). Corpora in teaching vocabulary and phraseology. In *The Routledge handbook of corpora and English language teaching and learning* (pp. 41–55). Routledge. <https://doi.org/10.4324/9781003002901-5>
- Titova, S., & Ignatova, S. D. (2024). The technology of application of multimodal linguistic corpora for foreign language interaction skill development. *Vestnik Tambovskogo Universiteta. Seriâ: Gumanitarnye Nauki*, 29(6), 1539–1549. <https://doi.org/10.20310/1810-0201-2024-29-6-1539-1549>
- Tracy-Ventura, N., & Braid, S. M. (2023). Corpus linguistics as a pedagogical tool. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley-Blackwell. <https://doi.org/10.1002/9781405198431.wbeal20475>
- Zaki, M. (2017). Corpus-based teaching in the Arabic classroom: Theoretical and practical perspectives. *International Journal of Applied Linguistics*, 27(2), 514–541. <https://doi.org/10.1111/IJAL.12159>
- Zeroual, I., El Kah, A., & Lakhouaja, A. (2018). Integrating corpus-based analyses in language teaching and learning: Challenges and guidelines. In Y. Tabii, M. Lazaar, M. Al Achhab, & N. Enneya (Eds.), *Big data, cloud and applications (Communications in Computer and Information Science, Vol. 872, pp. 534–545)*. Springer. https://doi.org/10.1007/978-3-319-96292-4_42