CORPUS-BASED DESIGN OF THE "SOCIAL MEDIA RESEARCH DICTIONARY" AT UNIDA GONTOR

Maulana Ashari[1], Yoke Suryadarma[2], Ahmad Ario Sofian[3]

[1, 2, 3] Universitas Darussalam Gontor, Indonesia

Corresponding E-mail: maulana.ashari@unida.gontor.ac.id

ABSTRACT

Corpus linguistics has become a key approach in language learning and lexicography, developing rapidly alongside digital technology in the Industry 4.0 era. However, its application in Arabic language education in Indonesia remains limited, especially in developing bilingual Indonesian–Arabic dictionaries for specialized domains. Addressing this gap, the present study designs and describes the development of the Indonesian–Arabic Dictionary of Social Media Research, a corpus-based lexicographic resource utilizing the SketchEngine platform for data processing, lexical extraction, and lemma selection. Using a qualitative descriptive method, the study applies corpus linguistics as both a methodological and analytical framework to ensure that dictionary entries represent authentic linguistic usage from academic and semi-academic Indonesian texts related to social media research. The results demonstrate that corpus-based analysis effectively identifies high-frequency vocabulary, collocations, and contextual word usage, producing entries that are linguistically representative and pedagogically meaningful. Expert validation further confirms the accuracy and usability of the compiled lemmas and their Arabic equivalents. This study concludes that corpus linguistics provides a systematic and data-driven foundation for bilingual dictionary development. The resulting dictionary contributes significantly to Arabic language learning in Indonesia and advances research in digital lexicography and corpus-based education.

Keywords: Arabic language learning, Corpus linguistics, Dictionary design, Social media research

ABSTRAK

Linguistik korpus telah menjadi pendekatan penting dalam pembelajaran bahasa dan leksikografi, yang berkembang pesat seiring dengan kemajuan teknologi digital pada era Industri 4.0. Namun demikian, penerapannya dalam pendidikan bahasa Arab di Indonesia masih terbatas, terutama dalam pengembangan kamus dwibahasa Indonesia–Arab untuk bidang-bidang khusus. Menanggapi kesenjangan tersebut, penelitian ini merancang dan mendeskripsikan pengembangan Kamus Indonesia–Arab Penelitian Media Sosial, yaitu sumber leksikografis berbasis korpus yang memanfaatkan platform SketchEngine untuk pemrosesan data, ekstraksi leksikal, dan seleksi lema. Dengan menggunakan metode deskriptif kualitatif, penelitian ini menerapkan linguistik korpus sebagai kerangka metodologis dan analitis guna memastikan bahwa entri kamus mencerminkan penggunaan bahasa yang autentik dari teks-teks akademik dan semi-akademik berbahasa Indonesia yang berkaitan dengan penelitian media sosial. Hasil penelitian menunjukkan bahwa analisis berbasis korpus secara efektif mengidentifikasi kosakata berfrekuensi tinggi, kolokasi, serta penggunaan kata dalam konteks yang tepat, sehingga menghasilkan entri yang representatif secara linguistik dan bermakna secara pedagogis. Validasi ahli lebih lanjut menegaskan ketepatan dan kebermanfaatan lema yang disusun beserta padanan bahasa Arabnya. Penelitian ini menyimpulkan bahwa linguistik korpus memberikan dasar yang sistematis dan berbasis data bagi pengembangan kamus dwibahasa. Kamus yang dihasilkan berkontribusi signifikan terhadap pembelajaran bahasa Arab di Indonesia serta memperkaya penelitian dalam bidang leksikografi digital dan pendidikan berbasis korpus.

Kata Kunci: Pembelajaran bahasa arab, Penelitian media sosial, Perancangan kamus, Linguistik corpus

INTRODUCTION

Learning Arabic as a foreign language requires effective supporting media that promote independent learning, particularly in enhancing vocabulary mastery (Hermawan, 2018). Among these media, bilingual dictionaries play a pivotal role as linguistic bridges that enable learners to access meanings, comprehend texts, and enrich their vocabulary when direct interaction with native speakers is limited (Alahmadi & Foltz, 2020; Barham & Clarke, 2022; Rahman et al., 2024). Numerous studies have confirmed that bilingual dictionaries significantly improve students' reading and writing skills (Suryadarma & Fakhiroh, 2020) and remain one of the most enduring and indispensable tools in language learning (Binti Jasni & Ardiansyah, 2020). Therefore, dictionaries are not merely supplementary aids but also foundational instruments for facilitating effective foreign language acquisition.

Despite their importance, preliminary observations conducted at the Department of Communication Science, University of Darussalam (UNIDA) Gontor, revealed that many students encounter substantial difficulties writing academic papers in Arabic, particularly in composing their undergraduate theses. These challenges stem primarily from lacking an Indonesian–Arabic bilingual dictionary specifically tailored to the academic and disciplinary context of communication and social media studies. Consequently, students face limitations in accurately expressing discipline-specific concepts in Arabic, which affects the quality of their academic writing and undermines the university's vision that emphasizes using Arabic as an educational language.

To address this issue, the present study proposes the development of a corpus-based Indonesian–Arabic bilingual dictionary specifically designed for social media research purposes. Corpus linguistics provides an empirically grounded approach that allows the extraction of representative and contextually relevant vocabulary through the analysis of word frequency and distribution (Gries, 2021; Hizbullah et al., 2020). This study employs blog entries and other digital linguistic resources as the foundational corpus for dictionary design. These electronic texts are systematically compiled to reflect authentic language use within social media research (Suryadarma & Fakhiroh, 2020).

Previous studies have demonstrated the significant potential of corpus-based dictionaries in producing more representative and authentic lexical entries (Almos et al., 2025). Rahimadinullah et al. emphasized that corpora provide empirical data for developing academic dictionaries (Rahimadinullah et al., 2023). Similarly, Rice and Zorn showed that corpus-driven approaches generate domain-sensitive vocabulary and sentiment values, outperforming general-purpose dictionaries (Rice & Zorn, 2021). Ivanov et al. introduced automated methods for constructing large corpus-based dictionaries through computational extraction and concreteness assessment (Ivanov & Solovyev, 2022). Abdelzaher further highlighted the role of corpus data in improving lexicographic decisions, including headword selection and collocation usage (Abdelzaher, 2022). These studies affirm that corpus-based lexicography offers greater accuracy, contextual relevance, and adaptability to contemporary linguistic developments, particularly in digital and media-related domains.

Nevertheless, a significant research gap remains: the lack of an Indonesian–Arabic bilingual dictionary specifically designed for academic and social media research contexts. Addressing this gap, the present study employs the SketchEngine tool to compile lemmas and entries derived from authentic linguistic data, ensuring that the resulting dictionary is theoretically sound, pedagogically, and practically applicable. The outcomes of this research are expected to contribute

to the advancement of bilingual lexicography and to support Arabic language learning among non-native speakers, particularly in academic writing and digital communication domains.

METHOD

This study employed a qualitative descriptive research design (Musthafa & Hermawan, 2018) aimed at systematically describing the process of designing a bilingual Indonesian–Arabic corpus-based dictionary within the domain of social media research. The qualitative descriptive approach was selected because it allows the researcher to document, explain, and interpret the sequential steps involved in dictionary construction, emphasizing linguistic and methodological rigor. The methodological framework was primarily based on Schierholz's lexicographic design model (Schierholz, 2015), which builds upon Wiegand's foundational theory of dictionary compilation (Adamska-Sałaciak, 2019). However, this study modified the framework to accommodate advances in corpus linguistics and digital lexicographic tools, thus aligning traditional lexicographic theory with modern computational practices.

The dictionary design process consisted of five interrelated stages: preparation, data and material collection, data analysis, expert evaluation and revision, and dictionary production. These stages were adapted from Schierholz's model and further refined according to Suryadarma and Fakhiroh to suit the bilingual and corpus-based nature of the project.

In the preparation stage, the researcher formulated the conceptual foundation of the dictionary, established its pedagogical and research purposes, and designed the initial template for dictionary entries. A detailed guidebook was drafted to standardize the format and ensure consistency in spelling conventions, transliteration, grammatical labeling, and semantic categorization. These preparatory efforts were essential to maintaining internal coherence and facilitating systematic data processing in the subsequent stages.

The data collection and analysis stage involved compiling a corpus of Indonesian-language materials related to social media research, drawn from linguistic blogs, peer-reviewed journal articles, and academic reports. Text selection adhered to three criteria: (a) written in formal or academic Indonesian, (b) containing terminology directly associated with social media studies, and (c) representing authentic and contextually relevant usage. The collected corpus was analyzed using SketchEngine, a state-of-the-art corpus analysis tool (Kilgarriff & Rundell, 2002), to identify frequency patterns, collocational structures, and contextual concordances. These analyses produced a list of candidate lemmas and corresponding usage contexts, which were then translated into Arabic equivalents accompanied by grammatical categories and example sentences. This stage ensured the dictionary entries were grounded in empirical linguistic data rather than intuition or prescriptive sources.

The evaluation and revision stage focused on expert validation to ensure the dictionary's academic reliability and pedagogical quality. Three experts, a professional lexicographer, an Arabic linguist, and a senior lecturer in Arabic language education, were selected based on their research experience, publication record, and involvement in Arabic teaching and lexicographic studies. Using a structured evaluation rubric, experts assessed three core aspects: (1) lemma accuracy and representativeness, (2) appropriateness of Arabic equivalents and grammatical labelling, and (3) pedagogical suitability for learners and researchers. Validation employed expert judgment and data triangulation to ensure analytical credibility, and the resulting inter-rater agreement (Cohen's Kappa = 0.87) indicated a high level of reliability. Expert feedback was systematically incorporated into the revision process, improving the clarity, consistency, and accuracy of entries.

Finally, the production stage involved the completed dictionary's formatting, printing, and intellectual property registration. This stage also included preparing a digital version for potential online dissemination to facilitate broader accessibility and academic use. By integrating Schierholz and Wiegand's theoretical lexicography framework with corpus-based linguistic methods, this research not only adheres to established dictionary design standards but also makes an original contribution to the field by demonstrating how corpus linguistics can enhance bilingual dictionary construction, particularly in specialized domains such as social media research.

RESULTS AND DISCUSSION
Result

After researching the design of this bilingual dictionary, the researcher found that the corpus-based linguistic approach is very useful for researchers in compiling or designing bilingual Indonesian-Arabic dictionaries. This is supported by Kennedy's finding that body language is the only source from which lexicographers can derive various types of information to present optimally in a dictionary (Kennedy, 2014). Additionally, this is evident from the dictionary compiled by Farouq, who succeeded in collecting 1489 Arabic vocabulary terms and their Indonesian equivalents in the field of Technology (Abdullah et al., 2024).

Hizbullah also confirmed this understanding, stating that one of the functions of using linguistic studies based on linguistic blogs (corpus linguistics) in Arabic language learning is to prepare a dictionary in specific fields(Hizbullah et al., 2020). One reason is that linguistic studies are based on linguistic codes, such as those used in agriculture, agricultural technology industries, or other fields. Therefore, this approach is well-suited for linguistics researchers or dictionary designers (Suryadarma & Fakhiroh, 2020).

Characteristics and Specifications of the Dictionary

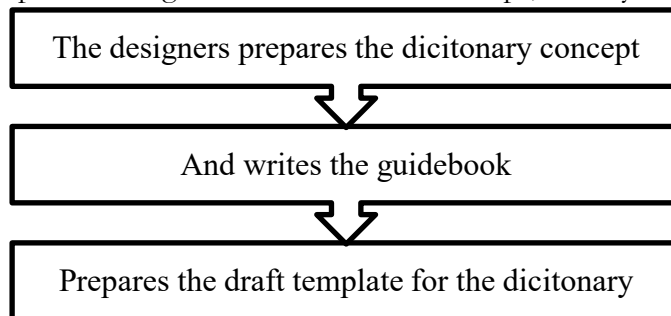This dictionary has several distinct features:

1. It is a bilingual dictionary whose source language is Indonesian, and the target language is Arabic (Suryadarma & Fakhiroh, 2020). The material in this dictionary consists of two languages, Indonesian (the source language) and Arabic (the target language).
2. The dictionary aims to expand linguistic information on Arabic vocabulary related to technology, industry, and agriculture. All entries (lemmas) in this dictionary are related to these fields.
3. The vocabulary in this dictionary consists of scientific terms and terminology associated with industrial agricultural technology. The collection method is based on linguistic blog entries (corpus linguistics), and the statements are available on the following pages.
4. The word order in this dictionary follows the alphabetical model, meaning that the entries are organized alphabetically according to the Indonesian alphabet (from A to Z), both horizontally and vertically.
5. The dictionary allows users to search for the meaning of words or vocabulary from the source language (Indonesian) to the target language (Arabic)(Zarkasyi & Yasin, 2015). The researcher then uses other dictionaries to obtain the meaning from the source language (Indonesian) to the target language (Arabic).

What Are the Steps in Designing this Dictionary?

Designing a dictionary requires a systematic process grounded in lexicography and linguistic research principles. Several recognized dictionary compilation approaches are influenced by theoretical perspectives and practical methodologies proposed by lexicographers (Suryadarma et al., 2019). One notable approach is that of Dr Ali Al Qasimi, who outlines five fundamental steps in dictionary creation: gathering relevant information and linguistic facts, selecting appropriate entries, arranging these entries according to a coherent and consistent system, writing and defining the material, and finally, compiling the completed work into a final product (Jamshidi et al., 2023). According to Abu Al-Faraj, the specific method employed by a lexicographer is less important than ensuring that the essential components of a dictionary are met. These components include determining the language of the dictionary, selecting and organizing the lexical material, and providing accurate and clear explanations of the words and their structures (al-Faraj & Aḥmad, 1966). As long as these requirements are fulfilled, the dictionary can be considered methodologically sound regardless of the design approach.

In designing the current dictionary, the researcher aims to provide Communication Science students with an accurate and comprehensive understanding of specialized Arabic terminology related to social media research. To achieve this, the researcher adopted the Wiegand method, as elaborated by Schierholz, which comprises five key steps in dictionary design and development. However, several modifications were made better to align the process with the objectives of this study. The adapted method incorporates additional steps for collecting linguistic data and analyzing it through the Arabic corpus linguistic approach (Suryadarma & Fakhiroh, 2020). This integration of corpus-based analysis ensures that the dictionary reflects authentic language use and contextually accurate meanings. The step-by-step procedure adopted in this research thus combines theoretical rigor with empirical data analysis to produce a specialized, user-oriented, and academically reliable dictionary tailored to Arabic communication and media studies.

1. Preparation stage. This includes several steps, namely:

```
┌─────────────────────────────────────────────┐
│   The designers prepares the dicitonary      │
│                 concept                       │
└─────────────────────────────────────────────┘
                      ↓
┌─────────────────────────────────────────────┐
│           And writes the guidebook           │
└─────────────────────────────────────────────┘
                      ↓
┌─────────────────────────────────────────────┐
│   Prepares the draft template for the        │
│              dicitonary                       │
└─────────────────────────────────────────────┘
```

The image illustrates the preparation stage in the process of designing a dictionary. This stage involves several essential steps that serve as the foundation before moving on to data collection or entry compilation. First, the designers prepare the dictionary concept, which includes defining the purpose of the dictionary, its type (e.g., bilingual or specialized), the target users, and the scope of the vocabulary to be included. This conceptual framework establishes the entire project's theoretical and practical direction. Second, the designers write the guidebook containing the technical and methodological guidelines for dictionary development. It includes instructions for entry writing, transliteration systems, definition formats, and consistency rules to ensure uniformity among all contributors. This guidebook is a reference manual for maintaining coherence throughout the dictionary-making process. Third, the designers prepare the draft template for the dictionary, which outlines the structure and layout of entries such as the lemma, meaning, example sentences, and target language equivalents, as well as the overall visual

presentation of the dictionary. This draft functions as a working model that will be refined and expanded later in the subsequent stages.

2. The stages of data and material collection and analysis, along with the steps carried out by the researcher in this stage, are as follows:
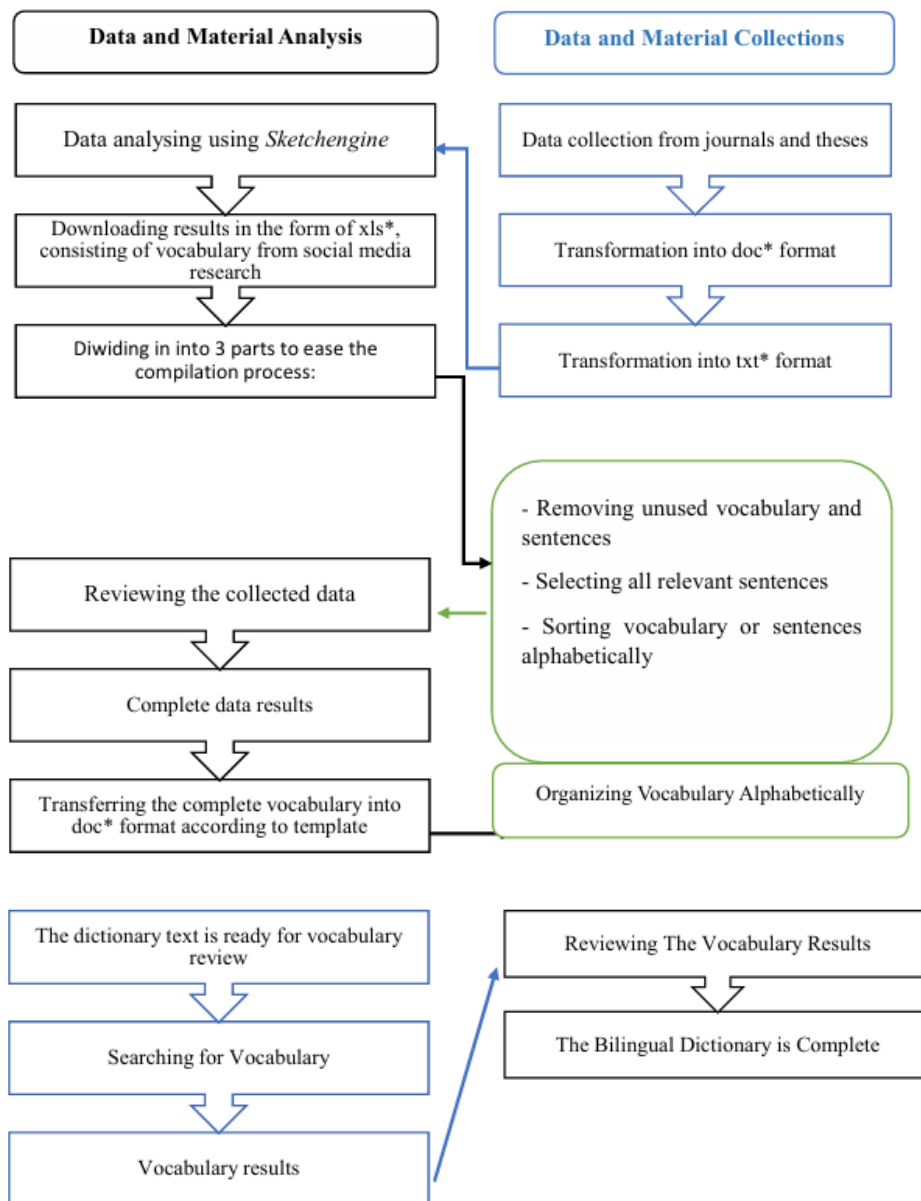
3. The stages of data and material evaluation, which consist of the following steps:
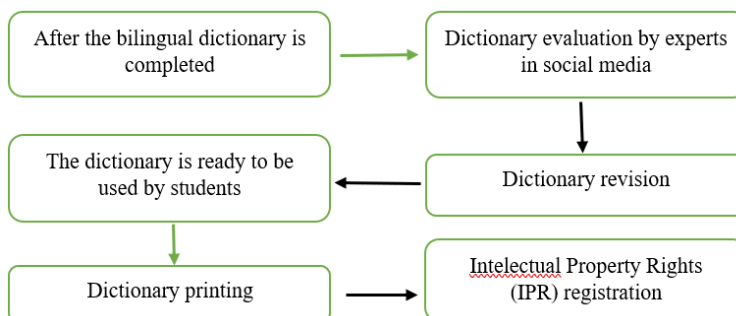
It is evident from this diagram that the researcher has followed this method, which is the most appropriate way to design a dictionary, as adopted in selecting vocabulary from Arabic-language blogs. Below is a detailed explanation of several key steps:

a) Data and Material Collection

The researcher sought to collect data or materials consisting of Indonesian scientific words or terms related to social media research. The researcher gathered materials from articles in print and online media (websites) to facilitate the search for these words or terms. The materials were sourced from printed documents and social media research. As for online media, data was taken from 45 scientific journal articles on social media research available at https://garuda.kemdikbud.go.id and theses by Communication Science students discussing social media.

b) Data Cleaning and Formatting

After collecting the data in the first step, the researcher inputs and selects it in a coding format* by removing all unnecessary elements (punctuation and numbers). The cleaned data resembles the following illustration:

Figure 1: Data transformation from Indonesian sentences or words into a doc* format

c) Convert the revised file containing Indonesian words or terms in doc* format to "Plain Text" (txt*) format with the "Unicode UTF-8" encoding, resulting in the following illustration:
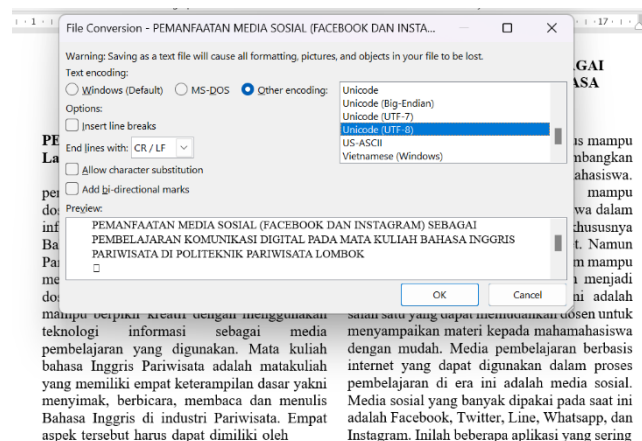

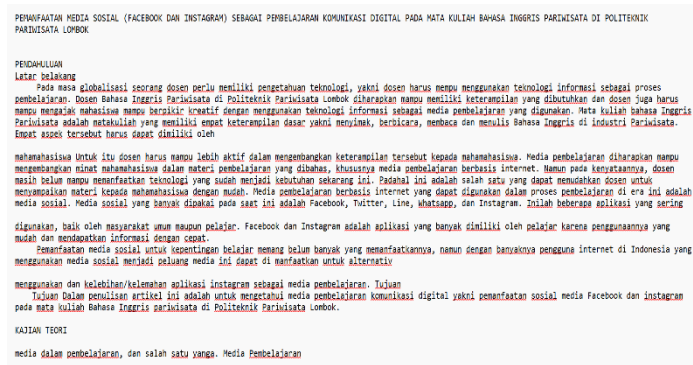**Figure 2: Transformation from doc* format to txt* format**

Figure 3: txt* format

d) Analyze data from the revised file in "Plain Text (txt*)" format based on linguistic codes using the web application SketchEngine through the website: https://www.sketchengine.eu, employing a wordlist. The description becomes clear as shown below:

Figure 4: Data Analysis using SketchEngine

e) After that, create a file in Excel format containing all vocabulary related to social media, arranged sequentially letter by letter. To clarify this statement, refer to the following illustration:

Figure 5: SketchEngine Output in xls* Format

f) Clean the data, focusing on basic vocabulary (root words), derivative vocabulary (derived words), and vocabulary with additional affixes (prefixes and suffixes) by removing unused words or vocabulary and eliminating all non-Indonesian words. This includes handling affixes to provide proper vocabulary or terms.

g) Review the data and transfer it to a doc* file format following the template prepared in the setup box.

h) The dictionary text is ready to search for synonyms from Indonesian to Arabic.

i) Search for synonyms from Indonesian to Arabic.

j) Evaluate the dictionary by lexicography and linguistic experts.

k) The dictionary is ready for printing.

From the data collection and processing based on linguistic blogs, the first data collection stage resulted in 2,753 vocabulary items related to social media research through the wordlist study using the SketchEngine application. After data editing, cleaning, and vocabulary addition, the second stage of data collection and processing yielded 8,333 items: root words and derivative words, including verbs, nouns, and adjectives related to social media research. These were added to the dictionary text in doc* format, and synonyms for the Indonesian terms were searched and matched to Arabic, following the dictionary preparation guidelines. For example:

A

abstrak	مُلَخَّصٌ
acuan	مَرْجَعٌ
adaptasi	التَكَيُّفُ
admin	إِدَارِي
administrasi	إِدَارَةٌ
adopsi	التَبَنِّي
afektif	وِجْدَانِي
agenda	جدول الأعمال
ahli	خَبِيرٌ

Figure 6: Vocabulary Structure in the Bilingual Dictionary

The research findings indicate that Corpus Linguistics entries can be used as an approach or foundation for compiling bilingual dictionaries in specific fields. Furthermore, creating a bilingual dictionary based on corpus linguistics has significant implications for helping researchers map and classify specific vocabulary in certain terms or fields systematically and easily. The most crucial stages in compiling a bilingual dictionary using linguistic blogs are selecting the data sources to be studied, processing the data within the linguistic blog application, and the researcher's ability to select vocabulary or lemmas (Entry/Lemma) to be included in the dictionary text. These three researcher capabilities are the primary requirements for compiling a dictionary based on linguistic plasticity (Corpus Linguistics).

Discussion

The findings of this study demonstrate that the corpus linguistic approach, implemented through the SketchEngine platform, is highly effective in designing the *Indonesian–Arabic Dictionary of Social Media Research*. The study successfully identified core vocabulary that accurately represents the language used in social media research by analyzing over eight thousand lexical items extracted from journal articles and theses. This result aligns with Tosun & Sofu, who emphasized that corpora are the most reliable linguistic source for lexicographers in producing accurate and contextually relevant dictionary entries (Tosun & Sofu, 2023). Similarly, Hizbullah et al.

confirmed that corpus-based approaches provide a solid foundation for constructing specialized dictionaries in various academic and professional domains (Hizbullah et al., 2020).

Comparatively, Abdullah, Sammah, and Rini (2024) designed a corpus-based Indonesian–Arabic bilingual dictionary in the technology field, while Suryadarma and Fakhiroh (2020) applied similar techniques in agricultural terminology. Although these studies demonstrated the feasibility of corpus-based lexicography in technical domains, the novelty of the present research lies in its focus on *social media research*, an emerging and interdisciplinary field that combines communication science, linguistics, and digital studies (Alenizi & Adawi, 2024; Hanks, 2012). This study thus extends the applicability of corpus linguistics beyond traditional domains such as technology and agriculture to communication and media, offering new perspectives for lexicographical innovation (Yao, 2019).

Theoretically, this research strengthens the framework of corpus-based bilingual lexicography by demonstrating that authentic linguistic data can be systematically transformed into structured and pedagogically useful lexical resources. It provides empirical evidence supporting the argument of Kilgarriff and Rundell (2002) that data-driven corpus methods lead to more representative lexical databases. Moreover, combining corpus analysis with bilingual dictionary design advances data-driven learning (DDL) principles in Arabic language education, as suggested by Mukhtar and Haroon, who highlighted the pedagogical value of corpus-informed resources in enhancing vocabulary acquisition.

The developed dictionary is a valuable academic tool for non-native Arabic learners, particularly students in Communication Science at UNIDA Gontor, to support thesis writing and academic publication. It bridges the terminological gap that often hinders comprehension of Arabic academic texts, thus improving Indonesian learners' academic literacy and translation competence. Furthermore, expert validation confirms the dictionary's linguistic accuracy and usability, reinforcing its potential as a reference model for other bilingual lexicographic projects in Indonesia.

However, this study also recognizes several limitations. The corpus was restricted to 45 Indonesian-language journal articles and selected student theses, limiting the lexical data's comprehensiveness. Future studies should expand the corpus by including a wider range of sources, such as international journal articles, digital discourse, and social media content, to enhance lexical diversity and coverage. Developing a digital or mobile-based dictionary version would also improve accessibility and interactivity, aligning with the goals of digital lexicography and modern corpus linguistics. In conclusion, this research provides both theoretical and practical contributions to the field of corpus-based Arabic lexicography. It demonstrates that integrating corpus linguistics into bilingual dictionary design ensures lexical accuracy and contextual authenticity and promotes the advancement of data-driven Arabic language learning in Indonesia.

CONCLUSION

This study concludes that the design of the Indonesian–Arabic Social Media Research Dictionary, developed through a corpus linguistic approach, has proven to be an effective and innovative model for bilingual lexicography. By utilizing the SketchEngine application, the research successfully identified and organized a systematic set of core vocabulary frequently used in social media research. This study's main contribution and novelty lie in its development of a specialized bilingual dictionary for a domain that has been largely unexplored in Indonesian–Arabic lexicographical studies, thereby bridging an important gap between linguistic theory and applied communication research.

From a practical perspective, this dictionary serves as a valuable linguistic resource for students, particularly those in the Communication Science Department at UNIDA Gontor, in addressing terminological and conceptual challenges encountered in writing academic articles and theses in Arabic. It also enhances Arabic academic literacy and promotes corpus-based lexicographical practices in Indonesia. However, the study acknowledges certain limitations, particularly regarding the restricted corpus size, which was limited to a selection of academic articles and theses. Therefore, future research should aim to expand the corpus by incorporating international journal publications, digital media, and other authentic linguistic sources. Furthermore, developing the dictionary into a digital or application-based format could significantly enhance its accessibility and usability. Such advancements would strengthen the foundation of Indonesian–Arabic bilingual lexicography and establish corpus linguistics as a robust methodological framework for future dictionary development across diverse academic and professional fields.

AUTHOR CONTRIBUTIONS STATEMENT

[MA] contributed to the conception and design of the study, conducted the data collection and analysis, and wrote the initial draft of the manuscript. [YS], as the primary supervisor, provided guidance throughout the research process, contributed to interpreting the results, and reviewed and revised the manuscript critically for important intellectual content. [AAS], as a dictionary designer, searched for synonyms for the dictionary, provided input, and helped refine the manuscript for publication. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENT

REFERENCES

Abdelzaher, E. M. (2022). An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years. *Lexikos*, *32*(1), 162–179. https://doi.org/10.5788/32-1-1714

Abdullah, A. F., Sammah, K., & Rini, F. S. (2024). Taṣmīm Qāmūsi at-Tiknūlūjiyā al-Mu'āṣir (Indūnīsiyā-'Arabī) li Iṡrā'i Mufrādāt Ladā Qismi Tiknūlūjiyā al-Ma'lūmāt bi Jāmi'ati Dārissalām Gontor. *Lisanudhad: Jurnal Bahasa, Pembelajaran, Dan Sastra Arab*, *11*(01), 111–128. https://doi.org/10.21111/lisanudhad.v11i01.11087

Adamska-Sałaciak, A. (2019). Lexicography and Theory: Clearing the Ground. *International Journal of Lexicography*, *32*(1), 1–19. https://doi.org/10.1093/ijl/ecy017

Alahmadi, A., & Foltz, A. (2020). Exploring the Effect of Lexical Inferencing and Dictionary Consultation on Undergraduate EFL Students' Vocabulary Acquisition. *PLOS ONE*, *15*(7), e0236798. https://doi.org/10.1371/journal.pone.0236798

Alenizi, A., & Adawi, R. (2024). Investigating the Effectiveness of Using Corpus-Based Developed Materials in Vocabulary Learning for Saudi EFL Students. *Forum for Linguistic Studies*, *6*(3), 721–745. https://doi.org/10.30564/fls.v6i3.6551

al-Faraj, A., & Aḥmad, M. (1966). *al-Maʿājim al-Lughawiyyah fī Ḍawʾ Dirāsāt ʿIlm al-Lughah al-Ḥadīth. A Study of the Arabic Language Dictionaries*.

Almos, R., Ermanto, E., & Ardi, H. (2025). Digital-Based Dictionary Compilation: Exploring Practical Steps, Technological Tools, and Pragmatic Analysis in Lexicography. *International Journal of Learning, Teaching and Educational Research*, *24*(3), 277–294. https://doi.org/10.26803/ijlter.24.3.13

Barham, K. A., & Clarke, R. (2022). Student-centered Experiences Using Dictionary Apps Within and Beyond the English Language Classroom in Palestine. *Sage Open*, *12*(4), 21582440221141697. https://doi.org/10.1177/21582440221141697

Binti Jasni, N. S., & Ardiansyah, A. A. (2020). Arabic Learning Based On E-Learning Using Arabic Teacher Website in the Pandemic Era of Covid-19. *Taʿlim al-ʾArabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, *4*(2), 122–137. https://doi.org/10.15575/jpba.v4i2.8601

Gries, S. Th. (2021). A New Approach to (Key) Keywords Analysis: Using Frequency, and Now Also Dispersion. *Research in Corpus Linguistics*, *9*(2), 1–33. https://doi.org/10.32714/ricl.09.02.02

Hanks, P. (2012). The Corpus Revolution in Lexicography. *International Journal of Lexicography*, *25*(4), 398–436. https://doi.org/10.1093/ijl/ecs026

Hermawan, A. (2018). *Metodologi Pembelajaran Bahasa Arab (Edisi Revisi)*. Remaja Rosdakarya.

Hizbullah, N., Arifa, Z., Suryadarma, Y., Hidayat, F., Muhyiddin, L., & Firmansyah, E. K. (2020). Source-Based Arabic Language Learning: A Corpus Linguistic Approach. *Humanities & Social Sciences Reviews*, *8*(3), 940–954. https://doi.org/10.18510/hssr.2020.8398

Ivanov, V., & Solovyev, V. (2022). Automatic Generation of a Large Dictionary with Concreteness/Abstractness Ratings Based on a Small Human Dictionary. *Journal of Intelligent & Fuzzy Systems*, *42*(5), 4513–4521. https://doi.org/10.3233/JIFS-219240

Jamshidi, F., Aghda, R. A., & Hammadi, M. (2023). *Methods of Expressing Apologies in Arabic from the Perspective of Social Linguistics*. *15*(1). https://doi.org/10.22108/RALL.2022.134371.1422

Kennedy, G. (2014). *An Introduction to Corpus Linguistics*. Routledge.

Kilgarriff, A., & Rundell, M. (2002). Lexical Profiling Software and its Lexicographic Applications-a Case Study. *Proc. EURALEX*, 807–818. Retrieved from https://euralex.org/elx_proceedings/Euralex2002/090_2002_V2_Adam%20Kilgarriff%20&%20Michael%20Rundell_Lexical%20Profiling%20Software%20and%20its%20Lexicographic%20Ap.pdf

Musthafa, I., & Hermawan, A. (2018). *Metodologi Penelitian Bahasa Arab: Konsep Dasar Strategi Metode Teknik*. Remaja Rosdakarya.

Rahimadinullah, M. Z. ʾAfafarrasyihab, Ahsanuddin, M., Irhamni, & Morhi, R. (2023). Development of Word Root-Based Digital Dictionary for Shorof Mastery. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, *5*(3), 265–284. https://doi.org/10.22219/jiz.v5i3.23024

Rahman, A., Ainin, M., Rosyidi, A. W., Bahy, Moh. B. A., Nasution, S., & Karnel, M. I. A. (2024). Enhancing Sharaf Mastery Through Ristek Muslim Digital Dictionary and Google Classroom Among Arabic Language Students. *Taʿlim al-ʾArabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, *8*(2), 1–15. https://doi.org/10.15575/jpba.v8i2.35971

Rice, D. R., & Zorn, C. (2021). Corpus-Based Dictionaries for Sentiment Analysis of Specialized Vocabularies. *Political Science Research and Methods*, *9*(1), 20–35. https://doi.org/10.1017/psrm.2019.10

Schierholz, S. J. (2015). Methods in Lexicography and Dictionary Research. *Lexikos*, *25*, 323–352. https://doi.org/10.5788/25-1-1302

Suryadarma, Y., Abdullah, A. F., & Rini, F. S. (2019). Khaṣāiṣ Qāmūs "Mu'jam Aṣ-Ṣiḥḥah" (Indūnīsī-'Arabī) lī Qism Aṣ-Ṣaydāliyyah wa Qism At-Tagẕiyyah Bi Jāmi'ati Dārussalam Gontor wa Muwāṣafātuhū. *LISANUDHAD*, *6*(2), 111–127. https://doi.org/10.21111/lisanudhad.v6i2.3521

Suryadarma, Y., & Fakhiroh, A. Z. (2020). Tashmīm Qāmus "al-Zirā'ah" Kawasīlah Ta'allum Al-'Arabiyyah Li Thalabah Qism al-Tiknūlūjiyā Al-Shinā'iyyah Al-Zirā'iyyah Muassasn 'Alā al-Mudawwanah al-Lughowiyyah. *LISANUDHAD*, *7*(2), 37–56. https://doi.org/10.21111/lisanudhad.v7i2.6744

Tosun, S., & Sofu, H. (2023). The Effectiveness Data-Driven Vocabulary Learning: Hands-on Concordancing through a Pedagogical Corpus. *Journal of Language and Education*, *9*(3), 176–190. https://doi.org/10.17323/jle.2023.12426

Yao, G. (2019). Vocabulary Learning through Data-Driven Learning in the Context of Spanish as a Foreign Language. *Research in Corpus Linguistics*, *7*(1), 18–46. https://doi.org/10.32714/ricl.07.02

Zarkasyi, A. H., & Yasin, R. F. (2015). *Mu'jam al-Mawrid li al-Duktūr Rūḥī al-Ba'labakkī (Dirāsah Waṣfiyyah Taḥlīliyyah)*. *Lisanudhad*, *2*(2), 83–96. Retrieved from https://ejournal.unida.gontor.ac.id/index.php/lisanu/article/view/648