International Journal of Learning, Teaching and Educational Research
Vol. 24, No. 3, pp. 277-294, March 2025
https://doi.org/10.26803/ijlter.24.3.13
Received Jan 18, 2025; Revised Mar 6, 2025; Accepted Mar 14, 2025

Digital-Based Dictionary Compilation: Exploring Practical Steps, Technological Tools, and Pragmatic Analysis in Lexicography

Rona Almos*, Ermanto, and Havid Ardi
Faculty of Language and Arts, State University of Padang, Padang, Indonesia

Abstract. This qualitative descriptive study investigated whether digital dictionary compilation can be systematically conducted using tools like Toolbox and ELAN. It focused on pragmatic approaches to improve accuracy while highlighting the challenges in transitioning from traditional to digital methods. This analysis revealed that the practice of compiling dictionaries can be digitally based. However, it emphasizes the potential threats of using digital technology in transitioning from conventional methods to a paperless culture required for technological development. This study used YouTube fragments as the main source of the research and note-taking as a collection method. The next step included identifying, classifying, and characterising vocabulary for digitally based dictionary input using these practical steps preparation, namely (1) searching for data in the form of attachments; 2) selecting and collecting data based on semantic meaning; 3) checking the prevalence of commonly used lemmas and sub-lemmas; 4) using the language corpus to filter words based on portraits of their usage; 5) collecting data based on lemmas' and sub-lemmas' usage habits; 6) Toolbox application setup; 7) preparing components in Toolbox; 8) inserting lemmas and sub-lemmas (vocabulary) into the Toolbox; 9) exporting dictionaries. Overall, this study expands the understanding of lexicography for students and highlights that lecturers need to adapt to often-overlooked technological changes. Future research should develop easy-to-use digital dictionary tools and training to help students and teachers adapt to new technologies while keeping useful traditional methods.

Keywords: dictionary compilation; digitally based; lexicography; Toolbox; ELAN; vocabulary

1. Introduction

The study of vocabulary is foundational to understanding language proficiency and its effective application in various contexts. Vocabulary reflects the quality

*Corresponding author: *Rona Almos, ronaalmos@student.unp.ac.id*

of a person's language skills and provides insights into linguistic development and communication capabilities (Dewandono, 2020). However, due to the lack of a universal understanding of the concepts underlying vocabulary and meaning, research on the practice of compiling dictionaries remains limited (Tarigan, 2009). This gap highlights the need for further exploration of theoretical and practical dimensions of lexicography, the branch of linguistics focused on the systematic study of words, their meanings, structures, and origins (Bergenholtz & Gouws, 2012; Klein, 2015).

Despite the growing body of research on vocabulary acquisition and lexical semantics, the practical implementation of lexicography in educational settings remains underexplored (Rizani, 2014). Many existing studies focus on vocabulary pedagogy, corpus linguistics, or computational language processing, yet they often overlook how students and educators can systematically compile dictionaries as part of the learning process (Milić et al., 2023; Morán, 2023; Moreno, 2023). A structured approach to dictionary compilation can enhance vocabulary retention and comprehension by enabling students to engage directly with linguistic data, categorize words based on semantic and pragmatic criteria, and develop a deeper understanding of lexical organization (Müller-Spitzer, 2023; Rahimadinullah et al., 2023; Shestakova & Kuleva, 2023). This perspective underscores the need to integrate digital tools and lexicographical methodologies into language education to bridge the gap between theory and practice.

Studying language skills and linguistic development is important, but the practical application of compiling dictionaries may not necessarily require a deep understanding of theoretical concepts. In addition, the value of lexicography may vary depending on the specific context or purpose of compiling a dictionary. For instance, in the field of computational lexicography, researchers often focus on developing algorithms and software to generate dictionaries for natural language processing systems (Martynova et al., 2015). In contrast, a lexicographer working on a historical dictionary must delve into etymology and semantic shifts over time to accurately document the evolution of words in a specific language (Cruse, 1995).

Given these diverse applications, the methodology used in lexicographic research must be adaptable to different linguistic objectives (Supriyanti, 2012). In educational contexts, digital lexicography tools such as ELAN and Toolbox provide opportunities for students to actively participate in dictionary compilation, moving beyond passive vocabulary learning. In contrast, lexicographers working in historical linguistics rely on extensive textual archives and diachronic linguistic analysis to trace word evolution. Meanwhile, computational lexicography integrates machine learning techniques to automate word sense disambiguation and dictionary generation. These varying approaches demonstrate that lexicography is a multidisciplinary field, requiring tailored methodologies based on the dictionary's intended use and target audience.

Existing studies in lexicography have predominantly addressed its theoretical underpinnings, including morphology, semantics, and phonology, but few have explored the pragmatic aspects of dictionary compilation (Jackson, 2002; Kridalaksana, 2003). Furthermore, with the rapid shift toward digital technologies, traditional methods of lexicography face challenges and opportunities. Digital tools such as Toolbox, ELAN applications, and language corpora offer innovative approaches to dictionary compilation but require structured methodologies to harness their potential effectively (Almos et al., 2023; Marliana, 2014).

Despite these advancements, significant gaps persist in the empirical and theoretical understanding of digital-based dictionary practices. Empirically, limited research explores how vocabulary is identified, classified, and contextualized in digital lexicographic practices (Dewi & Lestari, 2020). For instance, while widely used, platforms like Wiktionary lack the rigorous methodologies and quality control of traditional lexicography, often struggling to maintain semantic accuracy and represent localized linguistic variations (Collins, 2005). Theoretically, the transition from conventional to digital lexicography lacks comprehensive frameworks to address issues such as context-based pragmatic analysis, semantic accuracy, and integrating cultural nuances in dictionary entries (Bogoslovskayaa et al., 2015). These gaps underline the importance of further investigation, particularly in educational contexts, where students and lecturers are key stakeholders in adapting to technological shifts.

In addition to these challenges, the rapid evolution of digital tools necessitates continuous adaptation in lexicographic methodologies. Many existing digital dictionary platforms are designed primarily for general reference rather than for specialized educational purposes, making them less effective for structured vocabulary learning. Furthermore, while digital tools like ELAN and Toolbox offer advanced functionalities for data collection and classification, their adoption in academic settings is often hindered by a lack of comprehensive guidelines and training for users. To bridge these gaps, future research should focus on developing standardized lexicographic frameworks that integrate digital tools with pedagogical strategies, ensuring that dictionary compilation not only enhances linguistic knowledge but also aligns with broader educational objectives.

This study addressed these gaps by examining the practical steps of compiling digital-based dictionaries, particularly in higher education settings. This study focuses on higher education because university students have the necessary linguistic and analytical skills to engage in digital dictionary compilation. Unlike younger learners, they can critically analyze lexical structures and apply lexicographical methods systematically. In addition, higher education institutions emphasize digital literacy and corpus-based research, making digital lexicography highly relevant. Students also encounter specialized vocabulary in their fields, requiring structured dictionary compilation for academic support. Furthermore, universities provide expert supervision and institutional resources,

ensuring accuracy and validity in lexicographical work. This focus bridges the gap between traditional and digital lexicography, preparing students for future contributions to language research and technology. The findings emphasize the significance of contextual and pragmatic approaches in dictionary compilation while considering the challenges posed by transitioning to a paperless culture in the digital era.

2. Research Questions

This study sought to answer the following research questions:

1. What practical steps are involved in compiling a digitally based dictionary?
2. How can they be structured for educational purposes?
3. How can digital tools such as Toolbox, ELAN, and language corpora be effectively used in the lexicographic process?
4. What are the potential threats and challenges associated with transitioning from conventional lexicography to digital lexicographical practices in higher education?
5. How can context-based pragmatic analysis enhance the quality and relevance of digital dictionary entries?

3. Literature Review

The discussion on lexicography and lexicology presented in this section provides a foundational understanding of dictionary compilation and vocabulary studies. However, the explanation remains general and lacks a critical exploration of how lexicography has evolved in response to digital advancements. While references to Bergenholtz and Gouws (2012), Klein (2015), and Zygmunt (2019) establish the theoretical basis of lexicography, the literature would benefit from a discussion on computational lexicography and its implications in modern dictionary-making. In addition, Svensen's (2009) assertion that lexicography and lexicology are the same could be further scrutinized, as many scholars differentiate lexicology as a broader study of words and lexicography as its applied practice. Kridalaksana (2003) provides a structured definition of lexicology and lexicography, but integrating more recent studies on digital lexicography would strengthen the literature review's relevance in contemporary linguistic research.

The explanation of lexicography in this section correctly identifies its core activities, including researching, collecting, selecting, analyzing, and explaining lexical units. However, the description lacks depth in explaining how these processes are practically applied in modern lexicographic work, particularly with integrating digital tools. The mention of semantics as an essential aspect of lexicography is relevant, as meaning is a fundamental component in compiling dictionary entries. The reference to Cruse (1995) strengthens this argument by emphasizing the role of meaning analysis in lexicographic studies. However, the connection between semantics and lexicography could be further elaborated by discussing how pragmatic aspects and contextual variations influence dictionary definitions. The explanation of lemma and lexeme is conceptually accurate,

however, the discussion remains somewhat vague and should clarify the distinction between lemma as the base form of a word and lexeme as an abstract lexical unit that may have multiple word forms. In addition, it would be beneficial to explore how computational lexicography processes lemma and lexemes in digital dictionaries to reflect contemporary lexicographic advancements. Overall, while this section provides a fundamental overview of lexicography and its relation to semantics, expanding the discussion with practical applications and modern developments would enhance its comprehensiveness and relevance in the current linguistic landscape.

Cruse (1995) distinguishes the lexeme from the lexical unit. According to him, a lexicon is a lemma contained in a lexicon or "ideal dictionary" of a language. Lexical units are lexemes that are being analyzed as complex meaningful forms that have fixed characteristics, are related to other lexical units and are syntagmatically related to the context of speech. Therefore, in this study, the term 'lemma' is a unit of analysis which refers to semantics (Cruse, 1995). The meaning of a word is composed of some features, namely its relationship to the real world, the association that brings that relationship, its relationship with other words in the vocabulary, and the rules related to its relationship to other words in sentence and text structures (Jackson, 2002).

The discussion on the role of technology in dictionary compilation effectively acknowledges the growing significance of technological proficiency in modern lexicographic practices. The statement that technology enables the creation of dictionaries in line with contemporary demands (Sugianto et al., 2013; Utami & Dewi, 2020) is relevant, yet it lacks a detailed explanation of how these advancements specifically enhance lexicographic methodologies. While it is recognized that applications assist in transcribing and categorizing words, a more in-depth exploration of how different tools function and their impact on efficiency, accuracy, and user experience would strengthen the argument. The claim that technology aids researchers in grouping words based on meaning is valid but does not elaborate on whether such categorization is automated or requires manual intervention for accuracy.

Furthermore, the argument that applications in dictionary compilation practices increase student engagement (Dewi & Lestari, 2020; Nusi & Zaim, 2023; Puspitasari, 2019; Utami & Dewi, 2020; Wulandari et al., 2021) is well-founded. However, it would benefit from empirical evidence or case studies demonstrating measurable improvements in students' learning outcomes. While these studies suggest that technology makes lexicography more interactive, there is little discussion on potential drawbacks, such as students' dependency on technology or challenges adapting to new digital tools. In addition, issues of accessibility, particularly in regions with limited technological resources, should be considered to provide a more balanced perspective. Lastly, the discussion could be expanded by addressing the integration of artificial intelligence and corpus-based approaches in digital dictionary compilation. The contributions of machine learning in automating transcription and categorization would provide a broader perspective on how lexicography continues to evolve. Including these

aspects would create a more comprehensive and critical evaluation of the role of technology in modern dictionary development.

Therefore, the practice of compiling dictionaries based on lexicography studies needs to be carried out to help record and document endangered languages. Endangered languages have a very rich and unique vocabulary (Collins, 2005; Yance, 2017) but it is not well documented in the dictionary. By studying lexicology and lexicography, students can help preserve the diversity of languages and cultures (Marliana, 2014). This study not only contributes theoretically to understanding the practice of dictionary compilation based on lexicographic studies but also provides practical value by conceptually introducing new perspectives in lexicography research. This study identifies the practice of compiling dictionaries based on lexicography studies to implement dictionary compilation practices that can enrich students' teaching materials in higher education so that the learning process is more effective and efficient and does not deviate from the competencies that have been set (Erwinsyah, 2017; Iqbaluddin & Aisa, 2020; Syairi, 2013).

Research using dictionaries as lexicographic teaching materials was conducted by Bogoslovskayaa et al. (2015) and titled, "The name of the concept student in Russian and English languages: On lexicographical material." The results of their research showed that naming the concept of student in these two languages was not closely related to the Indo-European language family. Thus, the concept of student was considered in the semantic–cognitive aspect. Research conducted by Zainudin et al. (2013) showed that in lexicography courses, linguistics students are not only taught the principles of compiling dictionary entries but are encouraged to use the Malay corpus in the dictionary entry evaluation process. The use of corpus is key in modern lexicography (Zainudin et al., 2013). Therefore, students are introduced to the Malay corpus and Frame Semantics. Some in-depth findings suggest Frame Semantics as a suitable theory to assist lexicographers in comprehensively analysing bilingual dictionary entries. Practical and direct Frame Semantics analysis, using corpus as evidenced data, provides good training for linguistics students.

Similar research conducted by Almos et al. (2023) showed that the corpus can be used as a tool to support students in the learning process of lexicology and lexicography. By using the concordance feature that includes word combinations, learners can identify the differences between lemmas and sub-lemmas more clearly (Almos & Ladyanna, 2020). This concordance feature not only determines the class of words on a lemma but also gives the context needed to provide a deeper definition. In addition, the context in the concordance also helps students understand and define the lemma better. The research conducted by Smirnova (2015) discussed the case study of the lexeme bank showing the importance of considering the functional evaluative nature of language in compiling dictionary entries to achieve coherence of word meaning (Smirnova, 2015). This idea highlights the need to abandon the basic referential semantic principles and pay attention to the evaluative nature of human experience in lexical use in real texts.

4. Methodology

4.1. Research Design

The methodology employed in this study was grounded in a qualitative descriptive research design, which was well-suited for exploring the practice of dictionary compilation based on lexicography studies. Unlike quantitative research, which focuses on statistical hypothesis testing, qualitative descriptive research aims to describe and interpret findings to establish meaningful principles for future studies (Dewi & Lestari, 2020; Puspitasari, 2019; Wulandari et al., 2021). This approach was particularly relevant for examining contextual practices in dictionary preparation and their application in educational settings.

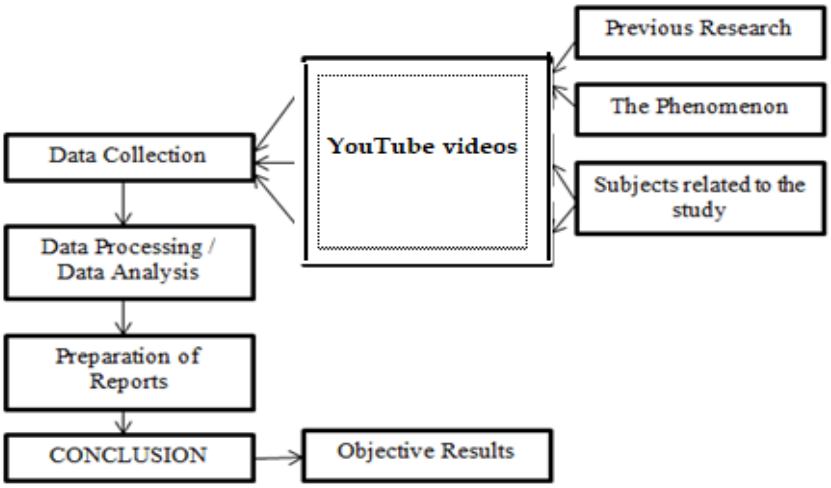

Figure 1: Scheme of Research Methodology

4.1. Source of the Data

The research's primary data comprised storyboard snippets from YouTube videos demonstrating dictionary compilation practices using the Toolbox and ELAN applications. YouTube videos were chosen as the primary data source because they provided authentic, diverse, and practical demonstrations of dictionary compilation using Toolbox and ELAN, allowing for real-world application analysis in a dynamic and accessible format. In addition, substantive data was sourced from online searches on lemma usage through www.google.co.id, providing a broad foundation for investigating lexicographic practices.

4.1. Data Collection Procedure

The data collection procedure involved several stages, beginning with the listening method, where relevant YouTube content was reviewed for examples of dictionary preparation practices. A note-taking technique was applied to capture significant information, particularly regarding the practical use of digital tools like Toolbox and ELAN. The collected data was then systematically identified, classified, and organized into typologies that highlighted distinct dictionary preparation practices. Data collection was concluded once sufficient examples were identified to address the research objectives. To ensure reliability

and validity, the categorized data underwent triangulation, where it was reviewed by relevant experts to minimize bias and provide a comprehensive understanding.

4.1. Data Analysis Procedure

The analysis of the collected data used the equivalent analysis method, with a specific focus on contextual equivalence. The process involved linking the data to the context of dictionary compilation to ensure that each lemma and sub-lemma aligns with its practical application. Interpretative analysis was then conducted on the classified and triangulated data, integrating linguistic intuition, scientific reasoning, and a literature review to establish accurate definitions and structures. The data was organized into tables to facilitate systematic comparisons, revealing patterns and relationships in dictionary preparation practices. The findings were further validated through expert triangulation, ensuring alignment with theoretical and practical lexicography. Special emphasis was placed on semantic and pragmatic accuracy, ensuring that the compiled dictionaries reflected linguistic precision and contextual relevance. Through this comprehensive methodology, the study rigorously examined the practical steps in digitally based dictionary compilation, contributing to advancements in educational practices and the adaptation of lexicographic methods to modern technological tools.

5. Results

5.1. Practical Steps for Compiling a Digitally Based Dictionary

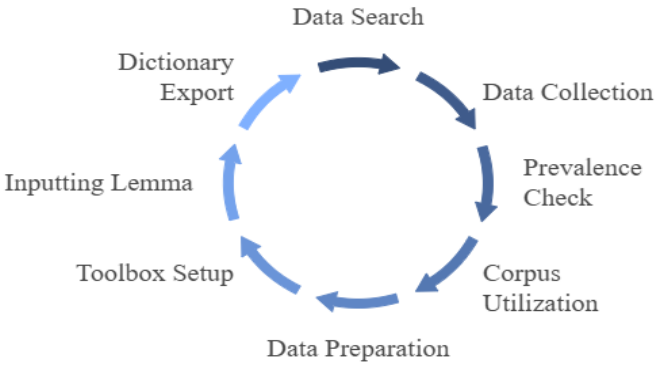

Figure 2: Practical Steps for Compiling a Digitally Based Dictionary

This study applied an established systematic process for compiling a digitally based dictionary and evaluated its practical implementation rather than defining new theoretical stages. The research presents empirical findings at each stage, highlighting key observations, challenges, and refinements necessary for effective dictionary compilation.

The first step involved searching for lemmas and sub-lemmas using Google, with advanced search techniques such as quotation marks (“ ”) and asterisks (*) to locate relevant data. The effectiveness of these techniques was tested by searching for commonly used words like technology, learning, and digital tools. Results showed that quotation marks helped retrieve exact phrases, while

asterisks expanded the search range. However, variations like technological vs. technology-related produced different results, indicating that search strategies impacted data consistency.

After data collection, words were classified based on their semantic meanings. For instance, in the context of educational technology, learning and training were initially grouped. However, the deeper analysis revealed that learning was a broader term while training referred to a structured process with specific objectives. This finding underscored the importance of precise classification criteria when grouping words in a digital dictionary. A prevalence check was then conducted to determine how frequently lemmas and sub-lemmas appeared in real-world contexts using Google search results. For example, e-learning appeared significantly more often than online learning materials, suggesting that e-learning was the dominant term in modern usage. This stage confirmed that some lemmas were more commonly used due to trends, user preferences, and regional differences.

To refine data accuracy, a linguistic corpus was used to analyze word usage across different contexts, such as academic articles, blogs, and social media. The term 'digital literacy,' for example, was found more frequently in academic publications, while tech-savviness was more common in informal blogs. This analysis demonstrated that corpus-based filtering helps refine dictionary entries to reflect real-world linguistic variations. However, challenges arose when implementing the data in the Toolbox application, particularly in assigning correct glosses and categorizing words by parts of speech. For instance, the word "access" functioned as both a noun (*internet access*) and a verb (*to access information*), requiring careful classification. Initial attempts led to incorrect tagging, which was later refined through iterative adjustments, illustrating the need for manual intervention to ensure accuracy in digital dictionary tools.

The final stage involved exporting the dictionary and validating its usability. Testing comparative linking techniques for contextual analysis revealed that words like "remote learning" and "distance education" appeared interchangeable in some contexts but had distinct meanings in others. This finding highlighted the importance of considering contextual equivalence when structuring dictionary entries. Overall, these findings confirmed that while the theoretical framework provided structured stages, practical application revealed variations and challenges. Differences in search strategies, classification inconsistencies, and context-dependent word usage emphasizes the need for iterative refinement in digital lexicography.

5.2. Effective Use of Digital Tools

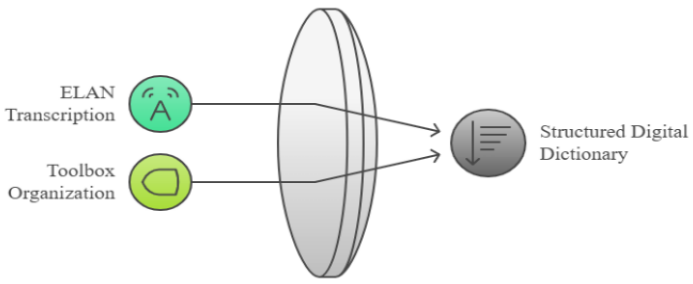

Figure 3: Effective Use of Digital Tools

Digital tools such as Toolbox and ELAN were crucial in streamlining the dictionary compilation process. ELAN was particularly useful for transcribing audio and video recordings into text, allowing the extraction of relevant linguistic data. For example, an audio recording of a speaker saying *"anak pisang di rimbo tak samo jo anak pisang di kandang"* was transcribed into segmented text, breaking it down into phonetic units and syntactic structures. This process facilitated the identification of key lexical items such as "*anak pisang*" (suckers of banana trees), "*rimbo*" (forest), and "*kendang*" (garden), which were later categorized based on their semantic roles. The analysis revealed that while "*anak pisang*" generally refers to banana shoots, its contextual meaning changes depending on location, implying the influence of the environment on plant growth.

Toolbox complemented this process by organizing and categorizing the extracted data. Once the transcribed words and phrases were imported into Toolbox, glosses were assigned to each lemma. For instance, the word "*gadang*" was entered with multiple annotations: as an adjective "*rumah gadang*" (*a traditional Minangkabau house*) and as an intensifier "*awak gadang urang dek kampuang*" (*an influential person in the village*). In addition, phonetic annotations were applied to words with complex pronunciations, such as "*caliak*" (/ʧa.li.ak/, meaning *to see*) and "*mancaliak*" (/man.ʧa.li.ak/, meaning *to observe carefully*). These entries were further analyzed based on their contextual usage, highlighting variations in meaning depending on formal or informal discourse.

The following table presents examples of data processed in ELAN and Toolbox.

Table 1: Examples of Data Input in ELAN and Toolbox

Word/Phrase	Phonetic Transcription	Part of Speech	Contextual Meaning	Gloss in Toolbox
Anak pisang	/a.nak pi.saŋ/	Noun	Banana shoot	A young banana plant
Anak pisang di rimbo	/a.nak pi.saŋ di rim.bo/	Phrase	Wild banana shoots grow differently	The environment affects plant growth
Gadang	/ga.daŋ/	Adjective	Large in size	Big, great
Gadang	/ga.daŋ/	Intensifier	Important or influential	A person of high status
Caliak	/ʧa.li.ak/	Verb	To see	To look at something
Mancaliak	/man.ʧa.li.ak/	Verb	To observe carefully	To examine closely

These entries illustrated how Minangkabau lexical items were processed using digital tools, ensuring that each word was accurately classified, transcribed, and defined. The findings demonstrated that ELAN and Toolbox provided a seamless workflow from raw linguistic data to a structured, exportable digital dictionary. Their ability to handle phonetic transcription, glossing, and semantic categorization ensured accuracy and efficiency in the dictionary compilation process. The user-friendly interface and compatibility with paperless workflows made these tools accessible for both novice and experienced lexicographers.

5.3. Potential Threats and Challenges in Digital Lexicography

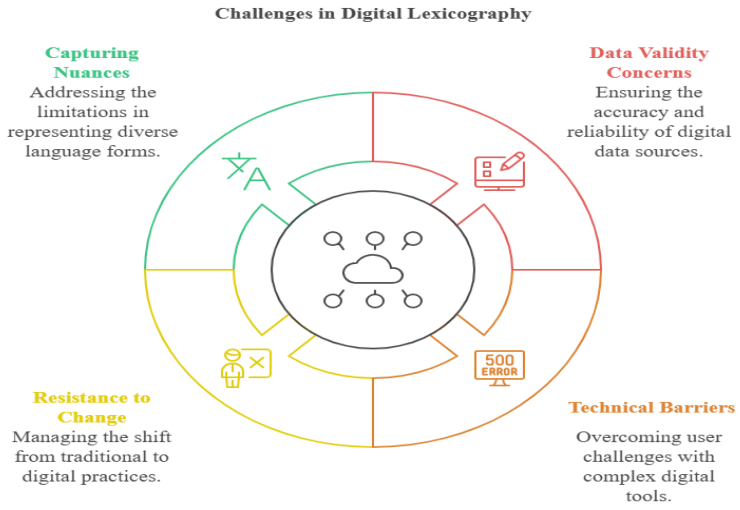

Figure 4: Potential Threats and Challenges in Digital Lexicography

Despite the benefits of transitioning to digital lexicographic practices, several challenges and threats must be addressed. First, the reliance on online sources like Google or language corpora led to concerns about the validity and reliability of the data. For instance, a prevalence check on the Minangkabau word "*mambasuik*" using different digital sources yielded inconsistent definitions. In some contexts, "*mambasuik*" meant "*mengeluarkan suara*" (to make a sound), while in others, it referred to "*berbicara pelan*" (to speak softly). Similarly, the term "*gulai*" was often translated as "*kari*" (curry) in Indonesian, but in Minangkabau, "*gulai*" refers to any dish cooked with spices, including those without coconut milk, making the translation misleading. A linguistic corpus contained outdated or contextually irrelevant usages, requiring expert validation. In this regard, triangulation with native speakers and language experts was essential to ensure that dictionary entries accurately represented linguistic reality. In our study, a set of Minangkabau lemmas and sub-lemmas were reviewed by experts to verify semantic accuracy, revealing that many entries required modifications due to inconsistencies in meaning or contextual usage.

Second, technical challenges associated with using applications like ELAN and Toolbox posed barriers for users who lacked adequate training. In one case,

students attempting to segment and annotate phonetic data for Minangkabau words like "*cabuik*" (*cabut, tarik secara cepat/to pull out quickly*) in ELAN faced difficulties aligning transcriptions with audio timestamps, resulting in errors in time-coded annotations. Similarly, users of Toolbox encountered challenges when entering glosses for Minangkabau terms that changed meaning depending on affixation, such as "*manampuah*" (*memukul dengan benda tumpul/to hit with a blunt object*) versus "*tanampuah*" (*terpukul dengan benda tumpul/to be hit with a blunt object*). These cases highlighted the need for structured training programs to ensure accurate annotation and categorization in digital lexicographic databases.

In addition, the shift from traditional paper-based lexicography to digital methods faced resistance from individuals accustomed to conventional practices. A survey conducted among experienced Minangkabau lexicographers showed that many still preferred manual compilations, citing concerns about digital reliability and usability. Implementing change management strategies, such as phased digital adoption and hybrid lexicographic workflows, could have eased this transition.

Lastly, digital tools struggled to capture nuanced language variations, such as dialects or informal expressions, potentially limiting the inclusivity of the dictionary. For example, an attempt to classify Minangkabau dialectal variations revealed that terms like "*caliak*" (*lihat/see*) and "*liek*" (*lihat dengan saksama/observe carefully*) often lacked direct equivalents in Indonesian, leading to potential misrepresentation. Similarly, informal expressions such as "*pai ka nan tarang*" (*pergi tanpa tujuan jelas/to wander aimlessly*) were not accurately categorized in standard lexicons. To address this, collaborative corpus-building efforts with native speakers enhanced the adaptability of digital dictionaries to linguistic diversity. These findings highlighted the importance of addressing technical, methodological, and linguistic challenges in digital lexicography to ensure accuracy, usability, and inclusivity.

5.4. Enhancing Dictionary Quality Through Context-Based Pragmatic Analysis

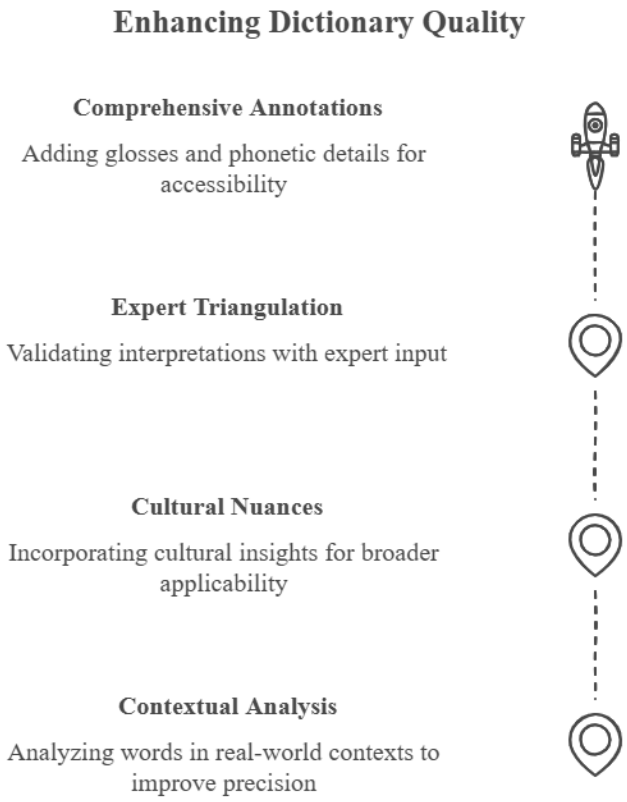

Figure 5: Enhancing Dictionary Quality through Context-Based Pragmatic Analysis

The study highlighted the importance of context-based pragmatic analysis in improving the quality and relevance of digital dictionary entries. The dictionary entries became more precise and meaningful by analyzing lemmas and sub-lemmas within their real-world contexts. This approach ensured that definitions aligned with both semantic and pragmatic aspects, reflecting actual usage in communication. Incorporating cultural and pragmatic nuances further enhanced the dictionary's applicability across diverse educational and linguistic contexts. Expert triangulation during the analysis phase validated interpretations, ensuring consistency with established linguistic principles. In addition, the inclusion of annotations such as glosses and phonetic details made the dictionary more comprehensive and accessible, catering to a wide range of users.

The effectiveness of context-based pragmatic analysis in lexicography was evident in words that had different meanings depending on context. In Minangkabau, the word "*manga*" could mean "*sudah*" (already) or "*mangga*" (mango) in Indonesian. For example, in "*Manga ka pai si Amak?*", "*manga*" means "*sudah*" (*Has Amak already left?*). However, in "*Manga ko rancak bana!*", "*manga*" functions as an intensifier, similar to "*memang*" in Indonesian (*Indeed, this is very beautiful!*). Without context, these differences were confusing. By adding contextual examples and gloss annotations in Toolbox, dictionary entries became clearer and more useful for learners and researchers.

Cultural and pragmatic factors also influenced word meanings, especially in translating expressions unique to Minangkabau. The word "*sarak*" could mean "*aturan adat*" (customary law) or "*larangan*" (prohibition), depending on usage. In "*Manuruik sarak, urang mudo indak boleh sakamar sebelum kawin*", "*sarak*" refers to customary law (*According to adat, young people could not share a room before marriage*). Meanwhile, in "*Jo sarak urang tuo, anak-anak ndak buliah bemain di jalan*", "*sarak*" means "*larangan*" (*According to the elders' rules, children were not allowed to play on the street*). Without proper annotation, these meanings could have been misunderstood in a dictionary. Including pragmatic annotations and real-world examples helps ensure that digital dictionaries accurately capture cultural and contextual meanings, making them more relevant for users.

5. Discussion

The practice of compiling dictionaries is one of the research subjects that has received considerable attention in the study of lexicography. Lexicography research seeks to understand the meaning of vocabulary in the form of phrases and sentences used in the community, including making word definitions, determining the correct use of words, and the practice of compiling word entries in dictionaries. Understanding the structure of language by expanding vocabulary in lexicography studies requires research, analysis of lexical data, and compiling dictionaries. This is necessary to document endangered languages that may not be recorded in dictionaries or encyclopedias to preserve the cultural and linguistic heritage and prevent the loss of knowledge (Almos et al., 2017; Almos & Ladyanna, 2019, 2020; Almos et al., 2023). For example, Shaniyazovna (2022) studied the practice of making dictionaries in the field of language documentation because dictionaries are the results of real research that

can be directly useful for the speaking community. Mastering the skills of how to compile dictionaries will be very useful in assisting researchers and staff in managing and accessing information related to their research (Lew, 2024). A comprehensive lexicography study requires a common understanding of the practice of compiling dictionaries (Marliana, 2014).

In the context of language, lexicology helps students understand the structure and evolution of vocabulary typical of their region, while lexicography allows them to compile dictionaries that reflect the richness of the local language and culture (Sigiyono, 2016; Yanti & Kurnia, 2016). Some higher education institutions have realised the importance of lexicography as an applied linguistic discipline, with its theory, methodology and practice, such as Rhodes University in South Africa (Nkomo, 2014), the University of Freiburg in Germany (Martynova et al., 2015) and the NoviSad University in Serbia (Prcic, 2020; Suntoko et al., 2022).

However, the digital practice of compiling dictionaries in lexicography is not the same as the practice of manually compiling dictionaries (Luan, 2022; Márkus & Dringó-Horváth, 2023). This lexicography study used the Toolbox and ELAN applications to facilitate the practice of compiling dictionaries by transcribing text in the form of audio and images to create an attractive design. This lexicographic study employed the Toolbox and ELAN applications to support dictionary compilation by transcribing text from audio and images, ensuring both linguistic accuracy and an effective presentation.

6. Conclusion

The integration of digital tools in dictionary compilation supports the growing need for technological advancements in lexicography. This study confirms that digital dictionary compilation enhances students' ability to understand vocabulary, phrases, and sentences in a structured manner. The use of ELAN facilitates accurate transcription of linguistic data from audio and video sources, while Toolbox enables systematic organization, annotation, and classification of lexical entries. These tools streamline dictionary development, making it more efficient and accessible in educational contexts.

However, the study also highlights several challenges in implementing digital tools for lexicography. While ELAN and Toolbox improve efficiency, issues such as data consistency, usability, and contextual accuracy must be addressed. The findings emphasize that digital dictionary compilation requires careful selection, classification, and validation of lexical data. Conducting prevalence checks, refining data through corpus analysis, and ensuring contextually relevant entries are essential to maintaining accuracy. Without these structured steps, inconsistencies in lemma classification and semantic representation may reduce the dictionary's reliability.

Moreover, the effectiveness of digital dictionary compilation depends on structured training and methodological guidance for both students and educators. Educators must integrate these tools into the learning process,

helping students not only acquire vocabulary knowledge but also develop analytical skills necessary for lexicographic research. Balancing technological advancements with traditional linguistic principles remains a key challenge.

Future research should explore advanced automation techniques, AI-driven lexicography, and user-friendly interfaces to enhance digital dictionary development further while maintaining linguistic accuracy and cultural relevance.

7. References

Almos, R., & Ladyanna, S. (2019). *Variety of rice-field lexicons from Luhak to Rantau*. https://eudl.eu/doi/10.4108/eai.23-3-2019.2284884

Almos, R., & Ladyanna, S. (2020). Lexicons classic of hunting in Minangkabau. In *Proceeding of The 13th International Conference onMalaysia-Indonesia Relations (PAHMI)* (pp. 47-51). https://www.researchgate.net/publication/338278674_Lexicons_Classic_of_Hunting_in_Minangkabau

Almos, R., Pramono, P., Hidayat, H. N., & Seswita, S. (2017). Teks klasik sebagai sumber pengembangan leksikografi. *Wacana Etnik*, *6*(2), 91–96. https://www.researchgate.net/publication/349984771_TEKS_KLASIK_SEBAGAI_SUMBER_PENGEMBANGAN_LEKSIKOGRAFI_MINANGKABAU

Almos, R., Pramono, P., Seswita, S., Asma, R. A., & Putri, N. O. (2023). Linguistik korpus: Sarana dan media pembelajaran pada mata kuliah leksikologi dan leksikografi di perguruan tinggi. *Lectura: Jurnal Pendidikan*, *14*(1), 45–59. https://doi.org/10.31849/lectura.v14i1.11705

Bergenholtz, H., & Gouws, R. H. (2012). What is lexicography? *Lexikos, 22*, 31–42. https://doi.org/10.5788/22-1-996

Bogoslovskayaa, Z. M., Novikovaa, V. S., & Itcenkoa, A. V. (2015). The name of the concept student in Russian and English languages: On lexicographical material. *International Conference for International Education and Cross-Cultural Communication. Problems and Solutions (IECC-2015)*, 301–305. https://doi.org/10.1016/j.sbspro.2015.11.653

Collins, J. T. (2005). Bahasa melayu bahasa dunia: Sejarah singkat. Yayasan Obor.

Cruse, D. A. (1995). *Lexical semantics*. Cambridge University Press.

Dewandono, W. A. (2020). Leksikologi dan leksikografi dalam pembuatan dan pemaknaan kamus. *Paramasastra*, 7(1), 16–26. http://dx.doi.org/10.26740/paramasastra.v7n1.p16

Dewi, M. S. A., & Lestari, N. A. P. (2020). E-modul interaktif berbasis proyek terhadap hasil belajar siswa. *Jurnal Ilmiah Pendidikan Dan Pembelajaran*, *4*(3), 433–441. https://doi.org/10.23887/jipp.v4i3.28035

Erwinsyah, A. (2017). Manajemen kelas dalam meningkatkan efektifitas proses belajar mengajar. *Tadbir: Jurnal Manajemen Pendidikan Islam*, *5*(2). https://journal.iaingorontalo.ac.id/index.php/tjmpi/article/view/392/309

Iqbaluddin, D., & Aisa, A. (2020). Asas pengembangan bahan ajar bahasa arab (Sosial budaya, psikologis, kebahasaan dan pendidikan). *El-Wasathiya: Jurnal Studi Agama*, 8(1). https://doi.org/10.35888/el-wasathiya.v8i1.3933

Jackson, H. (2002). *Lexicography: An introduction*. Routledge.

Klein, W. (2015). Lexicology and lexicography. *International Encyclopedia of the Social & Behavioral Sciences*, *13*(2), 937–941. https://doi.org/10.1016/b978-0-08-097086-8.53059-1

Kridalaksana, H. (2003). Sambutan ilmiah pusat leksikologi dan leksikografi fib ui dalam rintisan kajian leksikologi dan leksikografi. FIB UI.

Lew, R. (2024). Dictionaries and lexicography in the AI era. *Humanities and Social Sciences Communications,* 11(1). https://doi.org/10.1057/s41599-024-02889-7

Luan, Y. (2022). On dictionary innovation in the context of media convergence. *Cuadernos de Linguistica Hispanica*, *2022*(39). https://doi.org/10.19053/0121053X.n39.2022.13800

Márkus, K. P., & Dringó-Horváth, I. (2023). Developing dictionary skills through monolingual and bilingual English dictionaries at tertiary-level education in Hungary. *Lexikos*, *33*. https://doi.org/10.5788/33-1-1821

Marliana, L. (2014). Penyelesaian persoalan yang dihadapi mahasiswa dalam memahami leksikografi sebelum memasuki kerja menyusun kamus. In *Prosiding Seminar Internasional Kajian Leksikologi Dan Leksikografi Muthakhir*. https://lll.fib.ui.ac.id/wp-content/uploads/2015/03/Prosiding-Kajian-Leksikologi-dan-Leksikografi.pdf

Martynova, I. N., Shugayeva, N. Y., Kormilina, N. V., Nikitinskaya, L. V., Gromova, Y. N., & Vassilyeva, Y. N. (2015). Modern approaches to teaching English lexicography at the University of Freiburg, Germany. *Journal of Sustainable Development*, *8*(3), 220–226. https://doi.org/10.5539/jsd.v8n3p220

Milić, M., Kavgić, O. P., & Stojanović, A. K. (2023). A new English–Serbian dictionary of sports terms in the light of contemporary challenges. *Lexikos*, *33*. https://doi.org/10.5788/33-1-1796

Morán, M. T. F. (2023). The concept of lexicographic condensation: A review of and perspectives on digital lexicography. *Lexikos*, *33*. https://doi.org/10.5788/33-1-1824

Moreno, É. V. (2023). Towards a typological expansion of lexicographic works: Beyond the differentiation between printed dictionary and digital dictionary. *Revista de Linguistica y Lenguas Aplicadas*, *18*. https://doi.org/10.4995/rlyla.2023.18598

Müller-Spitzer, C. (2023). Gender stereotypes in dictionaries: The challenge of reconciling usage-based lexicography with the role of dictionaries as social agents. *Lexikos*, *33*(2). https://doi.org/10.5788/33-2-1843

Nkomo, D. (2014). Teaching lexicography at a South African university. *Per Linguam*, *30*(1), 55–70. https://doi.org/10.5785/30-1-558

Nusi, A., & Zaim, M. (2023). Philosophy of education in digital transformation: Ethical considerations for students' data security in online learning platforms. *Jurnal Ilmiah Pendidikan Scholastic*, *7*(3), 42–50. https://doi.org/10.36057/jips.v7i3.629

Prcic, T. (2020). Teaching lexicography as a university course: Theoretical, practical and critical considerations. *Lexikos*, *30*(1), 1–28. https://doi.org/10.5788/30-1-1597

Puspitasari, A. D. (2019). Penerapan media pembelajaran fisika menggunakan modul cetak dan model elektronik pada siswa SMA. *Jurnal Pendidikan Fisika*, *7*(1), 17–25. https://journal.uin-alauddin.ac.id/index.php/PendidikanFisika/article/view/7155

Rahimadinullah, M., Z., A., Ahsanuddin, A., Irhamnia, I., Morhi, R. (2022). Development of Word Root-Based Digital Dictionary of Shorof Mastery. *Izdihar : Journal of Arabic Language Teaching, Linguistics, and Literature, 5*(3), 265-284. https://doi.org/10.22219/jiz.v5i3.23034

Rizani, H. (2014). Kamus idris al-marbawi dalam tinjauan leksikologi (Analisis metode dan isi). *Alfaz, 2*(2), 224–237. https://media.neliti.com/media/publications/publications/233703-kamus-idris-al-marbawi-dalam-tinjauan-le-0250b208.pdf

Shaniyazovna, M. L. (2022). Lexicology and its main functions. *Texas Journal of Philology, Culture and History*, *5*. https://zienjournals.com/index.php/tjpch/article/view/1237

Shestakova, L. L., & Kuleva, A. S. (2023). "Dictionary of the language of Russian poetry (20th Century)" in the context of digital author dictionary. *Russkaya Rech*, *2023*(6). https://russkayarech.ru/en/archive/2023-6/63-76

Sigiyono. (2016). Menyusun kamus yang mencerminkan kekayaan bahasa dan budaya setempat. *Tantangan Leksikografis Bahasa-Bahasa Daerah Di Indonesia* (pp. 7–12). https://badanbahasa.kemdikbud.go.id/resource/doc/files/Prosiding_SLI_2016_compressed.pdf

Smirnova, A. Y. (2015). "Where is the bank?" or how to "find" different senses of a word. *Heliyon, 65*, 1–29. https://doi.org/10.1016/j.heliyon.2016.e00065

Sugianto, D., Abdullah, A. G., Elvyanti, S., & Muladi, U. (2013). Modul virtual: Multimedia flipbook dasar teknik digital. *Invotec*, *9*(2), 101–116. https://dx.doi.org/10.17509/invotec.v9i2.4860

Suntoko, O., Adham, M. J. I., & Ekawaty, N. (2022). Pelatihan cara menggunakan alat transkripsi elan. *Jurnal Pengabdian Kepada Masyarakat*, *1*(11), 3089–3094. https://doi.org/10.53625/jabdi.v1i11.1885

Supriyanti, N. (2012). *Praktik lesikografi atas nomina persona berorientasi gender dalam kamus besar bahasa indonesia* (4th ed.). Universitas Indonesia.

Svensen, B. (2009). *The handbook of lexicography, the theory and practice of dictionary-making*. https://doi.org/10.1093/ijl/ecq010

Syairi, K. A. (2013). Pengembangan bahan ajar bahasa Arab. *Jurnal Dinamika Ilmu*, *13*(1). https://www.researchgate.net/publication/307849604_Pengembangan_Bahan_Ajar_Bahasa_Arab

Tarigan, H. G. (2009). *Pengkajian pragmatik*. Angkasa.

Utami, Y. P., & Dewi, P. S. (2020). Model pembelajaran interaktif SPLDV dengan aplikasi rumah belajar. *Mathema*, *2*(1), 24–31. https://doi.org/10.33365/jm.v2i1.572

Wulandari, F., Yogica, R., & Darussyamsu, R. (2021). Analisis manfaat penggunaan e-modul interaktif sebagai media pembelajaran jarak jauh di masa pandemi Covid-19. *Khazanah Pendidikan*, *15*(2), 139–144. https://jurnalnasional.ump.ac.id/index.php/khazanah/article/view/10809

Yance, I. (2017). Situasi diglosia suku bonai di provinsi riau. *Madah, 8*(2), 127–136. https://madah.kemdikbud.go.id/index.php/madah/article/view/28

Yanti, R., & Kurnia, N. S. (2016). Menuju pemanfaatan leksikografis korpus bahasa daerah: Keanekaragaman tantangan dan peluang. In *Tantangan Leksikografis Bahasa-Bahasa Daerah Di Indonesia* (pp. 41–58). https://badanbahasa.kemdikbud.go.id/resource/doc/files/Prosiding_SLI_2016_compressed.pdf

Zahrah, H., Wargadinata, W., & Barry, N. A. (2021). Analisis E-dictionarry"Arab-Indonesia" yang tersedia di Play Store dengan pendekatan leksikolog. *Shaut Al-'Arabiyah, 9*(1). https://doi.org/10.24252/saa.v9i1.21494

Zainudin, I. S., Jalaluddin, N. H., & Bakar, K. T. A. (2013). The use of corpus and Frame Semantics in a lexicography class: Evaluating dictionary entries. *5th World Conference on Educational Sciences - WCES 2013* (pp. 2316–2320). https://doi.org/10.1016/j.sbspro.2014.01.566

Zygmunt, T. (2019). A review of "Selected chapters in English lexicology, Part I: Lexical semantics and lexicography" by Klaudia Bednarova-Gibova. *Language, Culture, Politics: International Journal*, *4*, 299–304. https://www.ceeol.com/search/article-detail?id=839853