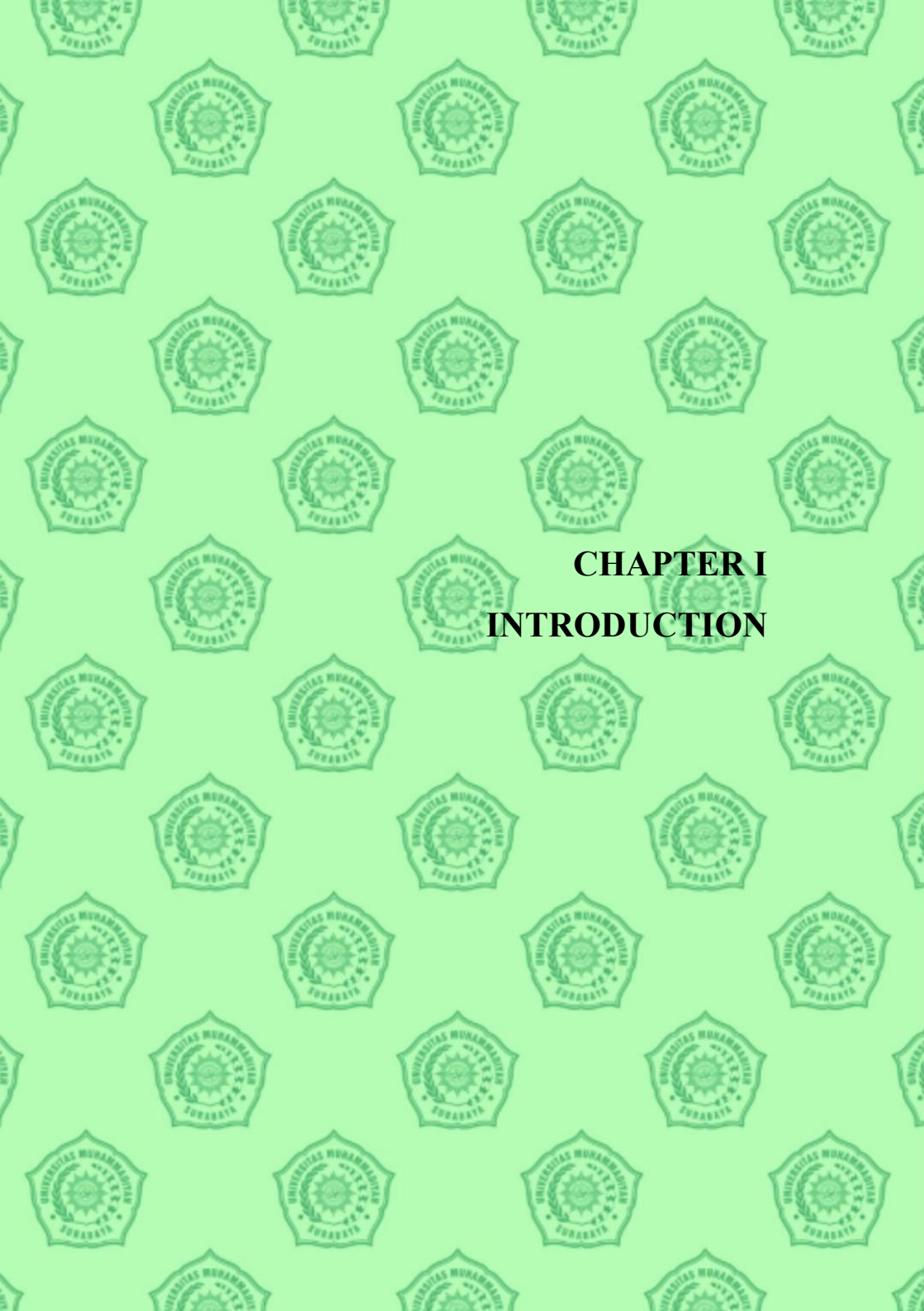


The background of the page is a repeating pattern of the Universitas Muhammadiyah Surabaya logo. The logo is a green pentagon with a central sun-like emblem, surrounded by the text 'UNIVERSITAS MUHAMMADIYAH SURABAYA'.

CHAPTER I

INTRODUCTION

CHAPTER I

INTRODUCTION

This chapter present some points of this part as the introduction in this thesis. The researcher organizes this chapter by dividing six parts; the first part is a background of research, the second is research questions, the third is research purpose, the fourth is scope and limitation, the fifth is a significance of the research, and last is the definition of key terms.

1.1 Background of the Study

Limited lexical variety is often found in the spoken English of EFL learners in international academic settings, which can lead to differences from native-like language use (Kolesnikova, 2016). However, spoken English remains an important tool for global academic communication, allowing multilingual students to take part in classroom discussions, presentations, and intercultural interaction. In Indonesia, English is mainly used as a foreign language, but it is increasingly present in higher education, especially in universities that host international students. In these learning environments, EFL students complete various speaking tasks while bringing vocabulary habits influenced by their first language into shared academic interactions.

Spoken English is a practical challenge for Cambodian students who are preparing to become future English teachers in the English Education program at Muhammadiyah University of Surabaya. Studying English in an EFL environment like Indonesia gives these students frequent exposure to local accents, learning practices, and interactions with peers from different backgrounds, which can influence the words they choose and how they express themselves. Previous research on Indonesian EFL learners also shows that limited vocabulary is a major barrier in speaking because students with restricted vocabularies often use many filler words and produce slow and simple sentences (Ali et al., 2025). The frequent use of the same common words is not always easy to notice when students speak in conversations. Because of this, the need for vocabulary development may not be clearly seen. Therefore, a corpus-

based frequency analysis of recorded student conversations can help identify words that appear repeatedly in the transcribed spoken data.

First-language influence may affect how Cambodian EFL learners choose and combine words when speaking English. Cambodian students are foreign EFL learners whose first language, Khmer, has structural differences from English, which may influence their vocabulary use in English communication. Previous studies support this idea, showing that first-language transfer can shape how learners use English vocabulary. For example, Rany et al. (2025) found that 88.45% of collocation errors in Cambodian students' writing were influenced by patterns from their first language. Therefore, analyzing lexical frequency and patterns in the spoken English of Cambodian students may provide useful insights into how their linguistic background influences vocabulary use in real communication.

Previous studies on Cambodian EFL learners mainly examine written language and learning factors rather than vocabulary use in spoken interaction. For example, some studies analyze collocation errors in Cambodian students' essay writing (Rany et al., 2025), while other research focuses on students' motivation and strategies in learning English (Heng, 2017). These studies provide useful insights into writing performance and learning attitudes, but they do not explain how Cambodian learners use vocabulary during informal spoken communication. In addition, although corpus methods are increasingly used in second language research, only a few studies have examined lexical patterns in learner spoken English in the Indonesian context (Ardi et al., 2023). Therefore, this study analyzes lexical frequency and lexical patterns in informal spoken English produced by Cambodian students at Muhammadiyah University of Surabaya using a corpus-based approach.

For example, (Hidayat & Al Viana, 2025) analyzed lexical patterns in Kapai-Kapai through a corpus approach with data processing using Antconc. Ro'ifah (2019) used corpus analysis to examine lexical coverage in ELT textbook reading materials, but the research focused on written instructional texts rather than spoken language produced by

learners. Likewise, Ozkan (2025) studied patterns and trends in modern English using large global corpora and discussed general language variation. The study did not specifically focus on learner speech in a particular local context. Therefore, research on lexical frequency and lexical patterns in informal spoken English produced by Cambodian students in Indonesia has received little attention, especially in topic-based interactions related to local contexts such as Surabaya's landmarks.

Corpus-based approaches offer a systematic way to examine vocabulary use and recurring expressions in learner language through frequency analysis and concordance tools (Poli & Cattolica, 2024; Sun & Park, 2023). In this study, LancesBox is used to analyze transcribed topic-based interactions and identify the dominant lexical choices in learner speech. This analysis helps show how EFL learners use vocabulary in real communication and how their lexical choices may differ from those of native speakers. By identifying these patterns, corpus analysis makes important features of learner language more visible for research and pedagogical evaluation. Therefore, the findings of this study are expected to provide context-specific insights that support more effective EFL teaching for international students.

Corpus linguistics provides a data-based approach to analyze authentic spoken language produced by learners (Kolesnikova, 2016). This approach helps researchers identify frequent lexical patterns that may not be visible through intuition-based analysis. By using corpus methods, researchers can examine word frequency and repeated expressions in learner speech to understand how vocabulary is used in real communication (Jaihow & Sinwongsuwat, 2024). Studies that analyze authentic student speech also show the common vocabulary and patterns that learners naturally produce during spoken interaction (Li et al., 2025).

Lexical patterns in spoken interaction should be examined through systematic study to better understand how learners use vocabulary in real communication (Jaihow & Sinwongsuwat, 2024; Kolesnikova, 2016). A corpus-based analysis that focuses on word frequency and recurring lexical patterns can provide concrete evidence of

how Cambodian students use language in informal spoken conversations. This type of corpus analysis helps move the discussion from general observations to a more empirical understanding of learner language use. Based on this reason, the present study develops specific research questions to investigate lexical frequency and patterns in Cambodian students' informal spoken English.

Only a small number of studies have examined the typical vocabulary patterns used by Cambodian EFL students in authentic spoken communication, especially in Indonesian university settings. By using corpus methods to analyze real spoken conversations, this study provides empirical evidence of how these learners use vocabulary during spontaneous interaction. The findings are expected to provide useful insights into students' word usage, which may contribute to teacher education and language teaching practices. In line with corpus pedagogy principles, identifying common lexical routines in spoken language helps teachers design instruction that better reflects students' real communication needs (Li et al., 2025). Moreover, Surabaya landmarks were selected as the topic because, based on the preliminary discussion, the students had already visited several landmarks in Surabaya.

1.2 Research Questions

This study is guided by the following research questions:

1. What are the most frequent lexical items produced by Cambodian students in topic-based informal spoken interaction about Surabaya landmarks?
2. How do recurring lexical patterns appear in the Cambodian students' informal spoken English corpus?

1.3 Research Objectives

Based on the research question, this study aims to:

1. To identify the most frequent lexical items produced by Cambodian students in topic-based informal spoken interaction about Surabaya landmarks.

2. To describe the recurring lexical patterns emerging from the frequency analysis of their spoken English corpus using LancsBox.

1.4 Significance of the Research

Some purposeful research is expected to have benefits on the audience of the research. It is also expected to benefit the writer or researcher himself. Accordingly, the writer wishes this to be significant from both theoretical and practical point of view.

1.4.1 Theoretical Significance

This research contributes to learner corpus linguistics by examining spoken lexical data from a learner group that has not been widely discussed in previous studies. It presents empirical evidence about word frequency and lexical patterns found in informal academic interactions produced by Cambodian EFL learners. Through a small-scale corpus approach, the study shows that quantitative corpus methods can be applied to analyze language use in a specific institutional context within Indonesian higher education. In this way, the study complements previous corpus research by highlighting local lexical patterns that appear in cross-national educational settings.

1.4.2 Practical Significance

Practically, this study gives English lecturers clear information about the common words used by Cambodian students in spoken English. This information can help lecturers design vocabulary activities that focus on the words students often use during speaking practice. The findings also provide useful evidence for the English Education Department to improve curriculum planning and strengthen support for international students. In addition, the research method used in this study can be applied in other small-scale corpus studies in different EFL programs.

Furthermore, the findings may also provide useful insights for instructors involved in Bahasa Indonesia for Foreign Speakers (BIPA) programs. Understanding how international students rely on high-frequency vocabulary and repetitive lexical patterns may help educators

design language-learning activities that better support vocabulary development and communicative competence among foreign learners in multilingual educational settings.

1.5 Scope and Limitation

1.5.1 Scope

The research analyzes lexical frequency and lexical patterns in a small spoken corpus produced by two Cambodian students from the English Education Department at Muhammadiyah University of Surabaya. Lexical frequency shows how often words appear in the students' speech, giving insight into the most commonly used vocabulary. Lexical patterns in corpus studies can take many forms, including n-grams, noun-based bundles, verb-based bundles, prepositional-based bundles, adjective-based bundles, and phraseological sequences. In this research, the focus is on verb-based bundles, which capture sequences of words containing verbs that students frequently use to describe activities, events, or landmarks in Surabaya. The data come from audio recordings of five topic-based conversations about Surabaya landmarks, with each conversation approximately 15–20 minutes long, and are transcribed using orthographic transcription. Analysis is carried out with LanesBox software to identify both word frequency and recurring verb-based bundles within the collected dataset.

1.5.2 Limitation

The research does not examine phoneme-level features or conversation analysis, such as turn-taking, interaction patterns, or discourse organization. It also does not analyze pronunciation, fluency, or overall communicative competence. This study is limited to morpheme-level analysis because the main focus is on vocabulary coverage and verb-based bundles used by Cambodian EFL students.

The analysis concentrates only on lexical features, especially verb-based bundles found in the spoken data. Since only two participants are involved, the findings describe specific patterns rather than general

patterns across a larger population. In addition, the conversations are topic-controlled, so the data reflect how students use vocabulary in guided discussions instead of fully spontaneous speech. Therefore, the conclusions provide descriptive insights into lexical frequency and verb-based bundles without making comparative or cause-effect claims. However, the findings may still be useful for lecturers to develop teaching strategies that focus on improving vocabulary use of Cambodian students, especially in similar learning contexts.

1.6 Definition of Key Terms

To avoid misunderstanding to understand this study, the definition of terms which are often found in this study as follows:

1. Corpus-Based Analysis

Corpus-based analysis is a research method used to study real language data collected in digital corpora. These corpora contain naturally occurring spoken or written texts produced by language users. Researchers use tools such as LanksBox to analyze the data and generate word frequency lists. The software also helps identify dominant lexical items and common patterns of language use. Through this process, researchers can observe authentic usage patterns that reflect how EFL learners actually produce language (Kolesnikova, 2016)

2. Lexical Frequency

Lexical frequency refers to the number of times specific words appear in a transcribed corpus dataset. This measurement helps researchers identify which vocabulary items learners use most often in spoken communication. By examining these frequencies, researchers can see patterns in learners' vocabulary choices. The results may also show repetition of certain words or limitations in lexical variety. Therefore, lexical frequency becomes an important indicator of vocabulary use among EFL learners (Jaihow & Sinwongsawat, 2024).

3. **Lexical Patterns**

Lexical patterns refer to repeated combinations of two or more words that appear regularly in a corpus. These patterns do not occur randomly but emerge through consistent use in language production. They include collocations and other multi-word expressions that learners use during speech. Studying these patterns helps researchers understand how learners organize vocabulary in real communication. This analysis also shows that language use often goes beyond individual words (Hidayat & Al Viana, 2025).

4. **Informal Spoken English**

Informal spoken English refers to spontaneous language used by EFL learners during everyday conversations. In academic settings, this type of speech often occurs in unstructured or lightly guided learning interactions. Learners usually rely on basic vocabulary, discourse markers, and repeated expressions during such communication. These features appear more frequently than complex structures that are typical in formal writing. As a result, informal spoken English reflects natural speaking behavior in real learning contexts (Umam, 2025).