



CHAPTER III

RESEARCH METHODOLOGY

CHAPTER III

RESEARCH METHODOLOGY

This chapter discusses the research methodology used in this study. It presents (1) Research Design, (2) Research Subject, (3) Data and Source of Data, (4) Data Collection, (5) Data Analysis.

3.1 Research Design

This study used a descriptive qualitative design within a corpus linguistics framework. It aimed to document lexical frequency and patterns in authentic spoken English data from EFL learners. Unlike experimental research, this method observed and reported natural language use without testing hypotheses or comparing groups. Corpus-based descriptive methods helped show how learners actually used words in speech (Kolesnikova, 2016).

The analysis focused on spoken English produced by two Cambodian students at Muhammadiyah University of Surabaya. Their recorded conversations provided spoken language data that could be examined to identify lexical frequency and recurring patterns. This methodological approach emphasized empirical description of language use based on corpus data. Such a design was appropriate for examining lexical features in a small set of learner speech within an academic environment.

3.2 Research Subject

The participants of this study were two Cambodian undergraduate students from the English Education Department at Muhammadiyah University of Surabaya. Both participants were 20 years old and were in their second semester at the time of data collection. Khmer was their first language, while English was learned as a foreign language. One participant had been learning English for approximately six years since the age of 14 and had only studied English through formal education, whereas the other had been learning English for about ten years

since primary school and had also attended an English course outside school. At the time of the study, both participants had been living and studying in Indonesia for approximately nine months. They reported using Khmer as their primary language in daily life while using English regularly for academic communication and interaction with friends.

These participants were considered appropriate for this study because they came from Cambodia and learned English in Indonesia, which is also a non-English-speaking country. This situation represents an English as a Foreign Language (EFL) learning context in which students acquire English outside English-speaking environments. Observing their language use provided insight into how international students use English in everyday communication within a foreign academic environment.

Their conversations about Surabaya's landmarks were recorded to collect natural and informal spoken English data. The topic of Surabaya's landmarks was selected because city landmarks are often among the first places visited by international students after arriving in a new city. Discussing these places encouraged the participants to share their experiences and observations, resulting in more natural informal conversations. The use of two participants allowed a focused analysis of lexical frequency and patterns in their spoken English.

3.3 Research Instrument

The main tool used in this study was LancsBox, a corpus analysis software designed to examine word frequency and lexical patterns. LancsBox could process transcribed spoken data and generate word frequency lists that showed how often each word appeared. Its Wordlist function helped researchers identify the most common words in small corpora (Jaihow & Sinwongsuwat, 2024). By using this software, counting became more objective and accurate, and it provided clear quantitative data about word use.

LancsBox was selected as the main tool in this study because it provided the GraphColl feature, which supported the analysis of lexical patterns in learner speech. This feature helped researchers visualize

collocations and see how words were connected in the dataset. In addition, LanksBox allowed researchers to identify recurring patterns through collocation analysis and showed which words often appeared together. Its interface supported both frequency counting and examination of words in context, which was important for descriptive corpus research. Therefore, LanksBox was an appropriate tool to ensure consistent and accurate extraction of lexical patterns from spoken learner data.

3.4 Data Collection

This study collected data through a topic-based semi-structured interview involving two Cambodian students. The interview consisted of 30 open-ended questions divided into three topics related to well-known landmarks in Surabaya: Heroes Monument, Suramadu Bridge, and Tunjungan Street. Each topic included ten questions that invited participants to describe their experiences, observations, opinions, and recommendations regarding the places they had visited.

The interview protocol was developed based on the objectives of the study to investigate lexical frequency and lexical patterns in informal spoken English. Open-ended questions were employed to encourage extended responses rather than simple yes-or-no answers, thereby generating a richer spoken corpus for analysis. The questions focused on personal experiences, physical descriptions of the landmarks, visitor activities, opinions, and recommendations. These aspects were selected to promote spontaneous spoken interaction within a familiar topic area. The three landmarks were chosen because the participants had previously visited them, allowing them to discuss the topics more comfortably and naturally during the interview sessions.

All interview sessions were audio-recorded with the participants' consent. After recording, the spoken data were transcribed orthographically into written form. The transcription included every word, repetition, and hesitation marker produced by the participants during the interaction. Fillers such as *umm*, *uh*, and *uhh* were retained because they reflected authentic spoken language production and provided evidence of real-time speech processing. The transcripts were

then converted into plain text (.txt) files for analysis in LancsBox. During corpus preparation, only technical transcription marks and irrelevant recording information were removed, while the original characteristics of the participants' speech were preserved.

3.5 Data Analysis

The main goal of this study is to identify core vocabulary patterns in students' speech using LancsBox.

1. Data analysis begins with frequency analysis through the Wordlist function, which generates a list of all words ranked by how often they appear.

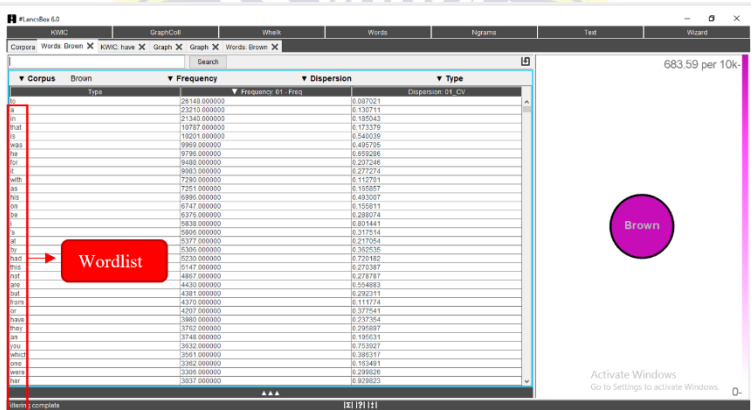


Figure 3.1 Wordlist Generated by Lancsbox

2. The second stage focuses on lexical patterns by using concordance analysis to identify word combinations that appear together frequently.

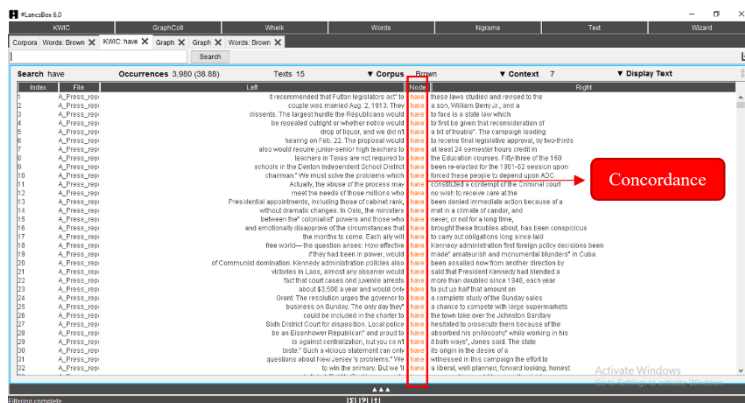


Figure 3.2 Concordance Results for the Word “have”

- The third stage involves collocation analysis, which examines how words often appear together more frequently than by chance. Collocation analysis helps reveal meaningful word pairings, common phrases, and lexical associations in students’ speech.

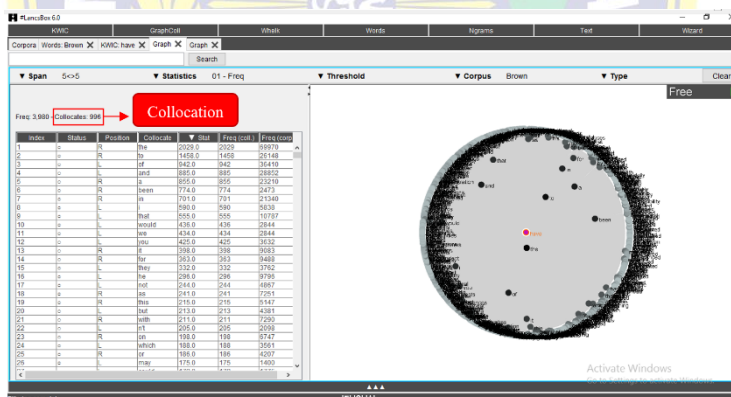


Figure 3.3 Collocation Result