



umsURA

Universitas Muhammadiyah Surabaya

SKRIPSI

**A CORPUS-BASED ANALYSIS OF LEXICAL
FREQUENCY AND PATTERNS IN INFORMAL
SPOKEN ENGLISH ABOUT SURABAYA'S
LANDMARKS BY CAMBODIAN STUDENTS
AT MUHAMMADIYAH UNIVERSITY OF
SURABAYA**

SRI WULANDARI
NIM. 20221111010

DOSEN PEMBIMBING
Sofi Yuniarti, S. S, M. Pd.
Sri Lestari, S. Pd., M. Hum.

**PROGRAM STUDI PENDIDIKAN BAHASA INGGRIS
FAKULTAS PENDIDIKAN, KOMUNIKASI, DAN SAINS
UNIVERSITAS MUHAMMADIYAH SURABAYA
2026**

**A CORPUS-BASED ANALYSIS OF LEXICAL
FREQUENCY AND PATTERNS IN INFORMAL
SPOKEN ENGLISH ABOUT SURABAYA'S
LANDMARKS BY CAMBODIAN STUDENTS AT
MUHAMMADIYAH UNIVERSITY OF SURABAYA**

UNDERGRADUATE THESIS

**Presented for Partial Fulfillment of the Requirements for
Undergraduate in English Education Department of
Education, Communication and Science Faculty**

SRI WULANDARI

NIM 20221111010

SUPERVISORS:

Sofi Yunianti, S. S, M. Pd.

Sri Lestari, S. Pd., M. Hum.

ENGLISH EDUCATION DEPARTMENT

FACULTY OF EDUCATION COMMUNICATION AND SCIENCE

MUHAMMADIYAH UNIVERSITY OF SURABAYA

2026

MOTTO AND DEDICATION

Motto :

"Allah did not bring you this far just to let you fail."

Dedication

With gratitude and humility, I dedicate this thesis to:

1. Allah SWT. The One who carried me through every difficulty, every doubt, and every prayer. This thesis is only possible because of His mercy, guidance, and countless blessings. Every step of this journey has been written by Him long before I could understand it. Alhamdulillah for bringing me this far. I pray that this thesis becomes beneficial knowledge, a source of goodness for others, and one of the righteous deeds that Allah accepts.

2. The Muslim Ummah and the People of Gaza. I dedicate this work as a small reminder that knowledge should always bring benefit to others. To our brothers and sisters in Gaza and to all Muslims facing hardship around the world, your patience, resilience, and unwavering faith have inspired countless hearts, including mine. This thesis may be a small contribution, but I sincerely pray that the knowledge I have gained will someday become beneficial for the Ummah. May Allah grant justice, peace, and the highest place in Jannah to those who have endured unimaginable trials.

3. My Parents. To my beloved mother, Suprihatin, and my beloved father, Gunadi, thank you for your endless love, prayers, sacrifices, and unwavering support. Thank you for believing in me even when I doubted myself. No words can fully express my gratitude for everything you have done, and no achievement could ever repay all the love you have given me. I hope I have made you proud, Mom and Dad.

Mom, I wish you were here to witness this moment. Although Allah had called you home before this day came, I believe your prayers are one of the reasons I made it this far. I hope this achievement becomes one of the prayers I can still send with love, and may Allah accept it as one of the endless rewards that continue to reach you. Thank you for raising me with so much love, kindness, and strength. Your presence continues to live in my heart every single day. I also remember my brother, whom Allah called home before I was born. Although we never met, I believe Allah had already written a special place for you in our family's story, and I pray that one day we will be reunited in Jannah.

Dad, thank you for every sacrifice, every prayer, and every silent effort that brought me to this point. Although you rarely express your feelings through words, I have always felt your love through your patience, responsibility, and unwavering support. Thank you for staying strong through every hardship and for never giving up on me. Your resilience has inspired me to keep moving forward, even during my most difficult days.

4. My Beloved Teacher. To my beloved teacher, Mam Ro'ifah, thank you for your kindness, sincerity, and unwavering support throughout my college journey. Although you were neither my thesis supervisor nor examiner, you always took the time to guide, encourage, and remind me whenever I needed it most. Your presence has been one of Allah's greatest blessings in my academic life, and your kindness will always remain in my heart. May Allah reward every lesson, prayer, and sincere support you have given me with abundant blessings.

5. My Future Family. To my future husband and children, wherever and whenever Allah has written our meeting, this thesis is a small part of the story before you arrived. I hope one day you will read these pages and know that every dream is worth pursuing with patience, effort, and trust in Allah's plan. May our future family always be built upon faith, gratitude, and the love of seeking knowledge together. Until that beautiful day comes, I entrust our story to the Best Planner, Allah SWT.

6. My Friends and Companions. To all of my friends from the English Education Department, Class of 2022, thank you for being part of one of the most meaningful chapters of my life. Especially to Beryl Imaneilla Mahsa, Nurul Ina Masruroh, Hana Furaidah, Virnanda Agustina Putri, and Regina Audi Yonrivania, thank you for every laugh, prayer, encouragement, and support throughout this journey. University life would not have been the same without all of you, and every memory we shared will always have a special place in my heart. May Allah bless each of your paths with success, happiness, and endless goodness.

7. Myself. Finally, I dedicate this thesis to myself. Thank you for not giving up, for surviving the difficult days, and for continuing to move forward even when the finish line seemed far away. Every challenge, revision, and sleepless night has shaped me into a stronger person than I ever imagined. Alhamdulillah, we made it.

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi yang ditulis oleh Sri Wulandari NIM 20221111010 dengan judul “A Corpus-Based Analysis of Lexical Frequency and Patterns in Informal Spoken English about Surabaya’s Landmarks by Cambodian Students at Muhammadiyah University of Surabaya” ini telah disetujui oleh dosen pembimbing untuk diujikan dalam UJIAN SKRIPSI pada tanggal 30 Juni 2026.

Dosen Pembimbing

Tanda Tangan

Tanggal

I. Sofi Yuniarti, S. S., M. Pd.



25 Juni 2026

II. Sri Lestari, S. Pd., M. Hum.



25 Juni 2026

Mengetahui
Ketua Program Studi Pendidikan Bahasa Inggris



Armeria Wijaya, S.S., M.Pd.

LEMBAR PENGESAHAN PANITIA UJIAN

Skripsi yang ditulis oleh **Sri Wulandari NIM 20221111010** dengan judul **“A Corpus-Based Analysis of Lexical Frequency and Patterns in Informal Spoken English about Surabaya’s Landmarks by Cambodian Students at Muhammadiyah University of Surabaya”** ini telah diuji dan dinyatakan sah oleh Panitia Ujian Tingkat Sarjana (S1) Fakultas Pendidikan Komunikasi dan Sains Universitas Muhammadiyah Surabaya sebagai salah satu syarat memperoleh gelar Sarjana Pendidikan pada tanggal 30 Juni 2026.

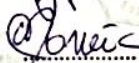
Dosen Penguji

I. Dr. Gusti Nur Hafifah., S. Pd.,
M. Pd.

II. Armeria Wijaya, S. S., M. Pd.

III. Sofi Yuniarti, S. S., M. Pd.

Tanda Tangan



Tanggal

30 Juni 2026

30 Juni 2026

30 Juni 2026

Mengetahui

Dekan Fakultas Pendidikan, Komunikasi, dan Sains



Achmad Hidayatullah, S. Pd., M. Pd., Ph. D.

PERNYATAAN TIDAK MELAKUKAN PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama : Sri Wulandari

NIM : 20221111010

Program Studi : Pendidikan Bahasa Inggris

Fakultas : Fakultas Pendidikan Komunikasi dan Sains

Menyatakan bahwa skripsi yang saya tulis ini benar-benar tulisan karya sendiri, bukan hasil plagiasi, baik sebagian maupun keseluruhan. Bila dikemudian hari terbukti hasil plagias, saya bersedia menerima sanksi akademik sesuai dengan ketentuan yang berlaku di Universitas Muhammadiyah Surabaya

Surabaya, 16 Juli 2026

Yang membuat pernyataan



Sri Wulndari
NIM. 20221111010

ABSTRACT

Wulandari, Sri. 2026 *A Corpus-Based Analysis of Lexical Frequency and Patterns in Informal Spoken English about Surabaya's Landmarks by Cambodian Students at Muhammadiyah University of Surabaya*. A Thesis English Education Department, Faculty of Education, Communication, and Science, Universitas Muhammadiyah Surabaya. Supervisor I: Sofi Yuniarti, S. S., M. Pd., Supervisor II: Sri Lestari, S.Pd., M. Hum.

Limited vocabulary and repetitive lexical patterns are common characteristics of English as a Foreign Language (EFL) learners' spoken production. However, studies examining lexical frequency and lexical patterns in the spoken English of Cambodian students in Indonesian higher education remain limited. This study aimed to identify the most frequent lexical items and describe the recurring lexical patterns produced by Cambodian students when speaking informally about Surabaya's landmarks. A descriptive qualitative design within a corpus linguistics framework was employed, involving two Cambodian students from the English Education Department at Muhammadiyah University of Surabaya. Transcribed recordings of topic-based informal conversations were analyzed using LancsBox through wordlist, concordance, and collocation analyses. The corpus consisted of 1,358 tokens and 252 lexical types, with nouns as the most dominant part of speech (38.89%) and A1-level CEFR vocabulary as the most frequent proficiency level (51.19%), reflecting only 12.45% coverage of the New General Service List. Wordlist analysis showed that function words such as *and*, *I*, *can*, and *the* occurred most frequently, while concordance analysis revealed simple Subject–Verb–Object patterns and recurring chunks such as *I think*, *I can*, and *people can*. Collocation analysis identified only one recurring pattern, very important, indicating limited collocational competence. These findings suggest that although the participants communicated effectively, their spoken English remained lexically limited and repetitive, offering practical insights for vocabulary-focused instruction in EFL contexts.

Keywords: *Corpus Linguistics, Lexical Frequency, Lexical Patterns*

ABSTRAK

Wulandari, Sri. 2026 *A Corpus-Based Analysis of Lexical Frequency and Patterns in Informal Spoken English about Surabaya's Landmarks by Cambodian Students at Muhammadiyah University of Surabaya*. Skripsi. Fakultas Pendidikan, Komunikasi, dan Sains. Universitas Muhammadiyah Surabaya. Pembimbing I: Sofi Yuniarti, S. S., M. Pd., Pembimbing II: Sri Lestari, S.Pd., M. Hum.

Kosakata yang terbatas dan pola leksikal yang berulang merupakan karakteristik umum dalam produksi tuturan pembelajar bahasa Inggris sebagai bahasa asing (EFL). Namun, penelitian yang mengkaji frekuensi dan pola leksikal dalam bahasa Inggris lisan mahasiswa Kamboja di perguruan tinggi Indonesia masih terbatas. Penelitian ini bertujuan untuk mengidentifikasi item leksikal yang paling sering digunakan serta mendeskripsikan pola leksikal yang berulang pada tuturan informal mahasiswa Kamboja ketika membicarakan landmark Kota Surabaya. Penelitian ini menggunakan desain deskriptif kualitatif dalam kerangka linguistik korpus, dengan melibatkan dua mahasiswa Kamboja dari Program Studi Pendidikan Bahasa Inggris Universitas Muhammadiyah Surabaya sebagai partisipan. Rekaman percakapan informal berbasis topik yang telah ditranskripsi dianalisis menggunakan LanksBox melalui analisis wordlist, concordance, dan collocation. Korpus penelitian terdiri atas 1.358 token dan 252 tipe leksikal, dengan nomina sebagai kelas kata yang paling dominan (38,89%) dan kosakata tingkat A1 CEFR sebagai tingkat kemahiran yang paling banyak muncul (51,19%), yang hanya mencakup 12,45% dari New General Service List. Analisis wordlist menunjukkan bahwa kata fungsi seperti *and*, *I*, *can*, dan *the* muncul paling sering, sementara analisis concordance menunjukkan pola kalimat sederhana Subjek–Verba–Objek serta ungkapan berulang seperti *I think*, *I can*, dan *people can*. Analisis collocation hanya mengidentifikasi satu kolokasi berulang, yaitu *very important*, yang menunjukkan keterbatasan kompetensi kolokasional. Temuan ini menunjukkan bahwa meskipun partisipan mampu berkomunikasi secara efektif, tuturan bahasa Inggris mereka tetap terbatas secara leksikal dan bersifat repetitif, sehingga memberikan wawasan praktis bagi pengajaran kosakata dalam konteks EFL.

Kata Kunci: *Frekuensi Leksikal, Linguistik Korpus, Pola Leksikal*

ACKNOWLEDGEMENT

Alhamdulillah Rabbil' Alamin, all praise and gratitude to Allah SWT for the presence of plenty of mercy, grace, strength, love, health, and guidance. Because of Him, the researcher has completed the studies for a bachelor's of Education degree at the English Education Department, Faculty of Teacher Training and Education, Universitas Muhammadiyah Surabaya, with the title **“A Corpus-Based Analysis of Lexical Frequency and Patterns in Informal Spoken English about Surabaya’s Landmarks by Cambodian Students at Muhammadiyah University of Surabaya”**

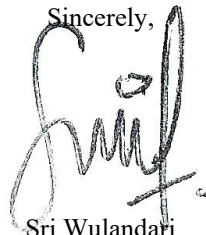
Peace and blessings to our Prophet Muhammad SAW, who has granted us passage from the era of ignorance to the age of enlightenment. Furthermore, the researcher has a lot of individuals by his side who help and inspire him to accomplish this thesis by offering advice, information, and motivation. The researcher so wishes to express gratitude to the following individuals:

1. Dr. Mundakir, S.Kep., Ns., M.Kep., the Rector of Universitas Muhammadiyah Surabaya
2. Achmad Hidayatullah, S.Pd., M.Pd., Ph.D. the Dean of Faculty of Education, Communication, and Science
3. Armeria Wijaya, S. S., M. Pd., as the Head of English Education Department, and also my first supervisor.
4. Sofi Yunianti, S. S., M. Pd. and Sri Lestari, S.Pd., M. Hum. as my supervisor who has guided and mentored me throughout the process of writing my thesis until its completion.
5. The entire English Education Department teachers which have guided and given me insight.
6. My parents, who supported, prayed, and guided me all the way to the finish line.
7. My classmates who have been with me throughout my college years.

The researcher acknowledges that this thesis is still far from flawless. Consequently, the researcher will be thrilled to embrace any helpful critiques and recommendations. Besides that, the researcher also hopes that this thesis will be useful for upcoming studies

Surabaya, 25 Juli 2026

Sincerely,

A handwritten signature in black ink, appearing to read 'Sri Wulandari', with a stylized flourish at the end.

Sri Wulandari

NIM. 20221111010

TABLE OF CONTENTS

MOTTO AND DEDICATION	i
LEMBAR PERSETUJUAN PEMBIMBING	iv
LEMBAR PENGESAHAN PANITIA UJIAN.....	v
PERNYATAAN TIDAK MELAKUKAN PLAGIASI	vi
ABSTRACT	vii
ABSTRAK.....	viii
ACKNOWLEDGEMENT	viii
TABLE OF CONTENTS	xi
LIST OF TABLES.....	xiv
LIST OF FIGURE.....	xv
LIST OF APPENDICES	xvi
CHAPTER I INTRODUCTION	1
1.1 Background of the Study.....	1
1.2 Research Questions	4
1.3 Research Objectives.....	4
1.4 Significance of the Research.....	5
1.4.1 Theoretical Significance	5
1.4.2 Practical Significance	5
1.5 Scope and Limitation	6
1.5.1 Scope.....	6
1.5.2 Limitation.....	6
1.6 Definition of Key Terms	7
CHAPTER II LITERATURE REVIEW	10
2.1 Corpus Linguistics in EFL Spoken Language Analysis.....	10

2.2 Vocabulary Coverage.....	11
2.3 Word Frequency in Spoken EFL Corpora.....	13
2.4 Lexical Patterns in Spoken EFL Corpora.....	14
2.4.1 Wordlist.....	15
2.4.2 Concordance.....	15
2.4.3 Collocation.....	16
2.5 Previous Studies.....	17
CHAPTER III RESEARCH METHODOLOGY.....	22
3.1 Research Design.....	22
3.2 Research Subject.....	22
3.3 Research Instrument.....	23
3.4 Data Collection	24
3.5 Data Analysis.....	25
CHAPTER IV FINDINGS AND DISCUSSION.....	28
4.1 Findings	28
4.1.1 NGSL Coverage Analysis.....	28
4.1.2 Part of Speech Distribution.....	28
4.1.3 CEFR Level Distribution.....	29
4.1.4 Lexical Frequency.....	30
4.2 Discussion	38
4.2.1 Lexical Item Frequency of Cambodian Student Spoken Interaction.....	38
4.2.2 Lexical Patterns of Cambodian Student Spoken Interaction	43
CHAPTER V CONCLUSION AND SUGGESTIONS	49
5.1 Conclusion	49
5.2 Suggestions	51

REFERENCES 54
APPENDICES..... 59

LIST OF TABLES

Table 4.1 Part of Speech Distribution	28
Table 4.2 CEFR Level Distribution.....	29
Table 4.3 Percentage of Category.....	30
Table 4.4 Top 1-50 Wordlist	30
Table 4.5 Collocation	38

LIST OF FIGURE

Figure 2.1 Display High-Frequency Vocabulary.....	12
Figure 3.1 Wordlist Generated by Lanksbox.....	25
Figure 3.2 Concordance Results for the Word “have”	26
Figure 3.3 Collocation Result.....	26
Figure 4.1 KWIC “and”	32
Figure 4.2 KWIC “I”	32
Figure 4.3 KWIC “can”	33
Figure 4.4 KWIC “the”	33
Figure 4.5 KWIC “have”	34
Figure 4.6 KWIC “in”	34
Figure 4.7 KWIC “is”	35
Figure 4.8 KWIC “because”	35
Figure 4.9 KWIC “people”	36
Figure 4.10 KWIC “a”	36

LIST OF APPENDICES

Appendix 1 : Transcribe Interview	59
Appendix 2 : Thesis Guidance Card.....	66
Appendix 3 : Plagiarism Check Result.....	67
Appendix 4: Letter of Plagiarism-Free.....	68
Appendix 5: Endorsement Letter.....	69
Appendix 6: The Researcher Biography	70