A CORPUS-BASED STUDY OF ENGLISH SYNONYMS: RESIDENT, INHABITANT, AND POPULATION

BY

NUNNAPUT PHUMEETHONG

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS IN ENGLISH LANGUAGE TEACHING

LANGUAGE INSTITUTE

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2024

A CORPUS-BASED STUDY OF ENGLISH SYNONYMS: RESIDENT, INHABITANT, AND POPULATION

BY

NUNNAPUT PHUMEETHONG

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS IN ENGLISH LANGUAGE TEACHING

LANGUAGE INSTITUTE

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2024

THAMMASAT UNIVERSITY
LANGUAGE INSTITUTE

AN INDEPENDENT STUDY

BY

NUNNAPUT PHUMEETHONG

ENTITLED

A CORPUS-BASED STUDY OF ENGLISH SYNONYMS:
RESIDENT, INHABITANT, AND POPULATION

was approved as partial fulfillment of the requirements for
the degree of Master of Arts in English Language Teaching

on July 16, 2025

Chairman Supakorn Phoocharoensil
(Associate Professor Supakorn Phoocharoensil, Ph.D.)

Member and Advisor
(Associate Professor Supong Tangkiengsirisin, Ph.D.)

Director Supakorn Phoocharoensil
(Associate Professor Supakorn Phoocharoensil, Ph.D.)

Independent Study Title	A CORPUS-BASED STUDY OF ENGLISH SYNONYMS: RESIDENT, INHABITANT, AND POPULATION
Author	Nunnaput Phumeethong
Degree	Master of Arts
Major Field/Faculty/University	English Language Teaching Language Institute Thammasat University
Independent Study Advisor	Associate Professor Supong Tangkiengsirisin, Ph.D.
Academic Year	2024

ABSTRACT

This corpus-based study investigates the distinctions between the near-synonymous nouns *resident*, *inhabitant*, and *population* in terms of frequency, formality, semantic preference, and collocations. Based on evidence from the Corpus of Contemporary American English (COCA), this study examines the distribution of three near-synonymous nouns across genres, their lexical patterns through collocational partners, and the semantic fields in which they typically occur. The findings reveal that *population* is the most frequent term, particularly dominant in institutional and academic genres, where it functions as an abstract, collective, and statistical noun. *Resident* occurs with moderate frequency and is widely used in semi-formal to formal contexts, especially in legal, civic, and medical discourse. *Inhabitant* is the least frequent of the three and appears primarily in formal, literary, or descriptive contexts—often relating to ethnicity, origin, or fictional settings. Collocational analysis shows that each word tends to pair with distinct sets of adjective and noun collocates, drawn from different semantic fields such as legality, continuity, geography, or vulnerability. These patterns suggest that while the three terms share a core denotative meaning, their usage is highly context-dependent and influenced by genre and register.

Keywords: corpus linguistics, frequency, formality, collocation, semantic preference, near-synonyms, COCA

ACKNOWLEDGEMENTS

I would like to extend my sincere appreciation to Associate Professor Dr. Supong Tangkiengsirisin, my advisor, for his invaluable guidance, constructive feedback, and unwavering support throughout the course of this Independent Study. His academic insight and professional expertise have been instrumental in shaping the direction and quality of this research. I also wish to express my gratitude to the faculty members and administrative staff of Language Institute of Thammasat University, for providing an intellectually stimulating environment and the necessary resources that facilitated the successful completion of this work. My heartfelt thanks also go to my family and close friends for their continuous encouragement, understanding, and moral support during the research and writing process. Lastly, I am indebted to all individuals who, in various ways, contributed to the accomplishment of this study. Their support is deeply appreciated.

Nunnaput Phumeethong

TABLE OF CONTENTS

	Page
ABSTRACT	(1)
ACKNOWLEDGEMENTS	(3)
LIST OF TABLES	(7)
CHAPTER 1 INTRODUCTION	1
1.1 Background of the study	1
1.2 Research objectives	2
1.3 Research questions	3
1.4 Definitions of key terms	3
1.5 Scope of the study	4
1.6 Significance of the study	4
CHAPTER 2 REVIEW OF LITERATURE	5
2.1 Introduction	5
2.2 The concept of synonymy in linguistics	5
2.2.1 The definitions of synonyms	5
2.2.2 Absolute synonyms, strict synonyms and loose synonyms	6
2.3 Criteria for distinguishing near-synonyms	8
2.3.1 Connotation and emotive tone	8
2.3.2 Degree of formality (Register)	9
2.3.3 Contextual appropriateness and register	10
2.3.4 Vocabulary differences across dialects	12
2.3.5 Spelling variations between British and American English	12
2.3.6 Collocation	13

2.3.6.1 Types of collocations 13
2.3.6.2 Collocational differences between synonyms 14
2.3.7 Grammatical patterns 14
2.3.7.1 Transitivity and verb agreement structures 15
2.3.7.2 Prepositional usage with synonyms 15
2.3.7.3 Passive vs. active voice constraints 16
2.4 Corpus linguistics and word studies 16
2.4.1 Concordance 17
2.4.2 Collocation 18
2.4.3 Semantic preference 18
2.5 Previous studies 20

CHAPTER 3 RESEARCH METHODOLOGY 28

3.1 Introduction 28
3.2 Selected words and justification 28
3.3 Data collection 29
3.3.1 Research instruments 29
3.3.1.1 Corpus of Contemporary American English (COCA) 29
3.3.2 Corpus selection 29
3.3.3 Data extraction 29
3.3.4 Data analysis 30
3.3.4.1 Frequency analysis 30
3.3.4.2 Collocational analysis 31
3.3.4.3 Semantic preference analysis 31

CHAPTER 4 FINDINGS 32

4.1 Frequency and formality 32
4.2 Collocational pattern 34

4.3 Semantic preference 37
4.3.1 Semantic preference of *resident* 37
4.3.2 Semantic preference of *inhabitant* 40
4.3.3 Semantic preference of *population* 42
4.4 Comparison of themes between *resident*, *inhabitant*, and *population* 44

CHAPTER 5 DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS 47

5.1 Summary of the findings 47
5.1.1 Frequency and distribution across genres 47
5.1.2 Collocational patterns 48
5.1.3 Semantic preference 48
5.2 Conclusion 49
5.3 Pedagogical implications 50
5.4 Limitations 51

REFERENCES 53

LIST OF TABLES

Tables	Page
1 Overall frequency and distribution of *Resident, Inhabitant and Population* across the eight genres from COCA	32
2 Adjective collocations of *Resident, Inhabitant, and Population* from COCA	34
3 Semantic preference of *Resident*	37
4 Semantic preference of *Inhabitant*	40
5 Semantic preference of *Population*	42
6 Comparison of themes between *Resident, Inhabitant, and Population*	44

CHAPTER 1
INTRODUCTION

1.1 Background of the Study

As English has become a global lingua franca—facilitating international communication, business, and education worldwide—the study of vocabulary is undeniably a vital aspect of learning the language. A strong vocabulary enables speakers to express their ideas with precision and clarity. Synonymy, or relation of a word with similar senses, is a persistent issue in vocabulary studies for learners and linguists. While dictionaries often add words as synonyms, in reality, their occurrence in spontaneous usage follows intricate collocation rules of co-occurrence, grammatical constructions, register, and shades of meaning (Palmer, 1997; Murphy, 2003). Synonymy is of most concern in the context of English as a Second Language (ESL) teaching, where students must make distinctions among words with roughly similar meaning but differing uses in context.

Misuse of near-synonyms can lead to awkward phrases, unidiomaticity, and unintended meanings in sentences. One of the most frequent misuses in near-synonyms is the use of the words *resident, inhabitant,* and *population,* which all mean persons or groups residing in a place. However, these words vary in their semantic constraints, syntactic distribution, and collocational patterns. Also, these sensitive nuances are not always treated in a straightforward way in traditional dictionaries and ESL course books, which do tend to provide lists of synonyms without sufficient context. Students have to rely on intuition, which is why mistakes come about with such frequency. One of the most important aspects of learning the English language is the correct use of near-synonyms.

Although dictionaries define resident, inhabitant, and population as habitation and demographic classification synonyms, they are actually used very differently. Students often assume these words are interchangeable, leading to errors such as *"The city's population is predominantly middle-class residents,"* when it should be *"The city's population is predominantly middle-class inhabitants."* Another example is *"She is a population of this town,"* which should be corrected to *"She is a resident of this town."*

These mistakes stem from a lack of understanding regarding the grammatical constraints, collocational usage, and subtle semantic differences among these near-synonyms. Previous research has confirmed that learners have difficulty handling near-synonyms owing to poor dictionary definitions and insufficient proper exposure to original use conventions (Phoocharoensil, 2010; Xiao & McEnery, 2006). Even though some research has explored synonymy from corpus linguistics, the existing literature includes minimal research on these specific words. The words resident, inhabitant, and population were selected for study since they are a highly relevant and frequently conflated set of near-synonyms in English. They are applied across a wide variety of registers, particularly in urban studies, legal claims, and research into populations. Though similar, they differ in terms of grammatical structure, collocation, and register-specific usage.

Other synonym pairs have been contrasted in previous research, yet little has been done comparing these three words in a systematic way based on corpus data. By filling this gap, the way they are used will be understood better, and it will be possible to facilitate significant insights for English language learners, teachers, and lexicographers. In order to achieve this, corpus linguistics has emerged as a vital tool in examining how synonyms function in authentic language use. With the analysis of large, actual datasets such as the Corpus of Contemporary American English (COCA), researchers can search for patterns of collocation, register usage, and grammatical composition that distinguish near-synonyms. In this study, the near-synonyms resident, inhabitant, and population are contrasted with each other across various registers, their grammatical composition analyzed, and their patterns of collocation looked for. The findings will provide insights into the usage of these words in real-life language contexts and offer practical recommendations for ESL learners and instructors.

1.2 Research Objectives

This study aims to address the aforementioned problem by conducting an in-depth corpus-based analysis of the near-synonyms *resident, inhabitant,* and *population.* The specific objectives of this research are:

1. To investigate the frequency patterns of *resident, inhabitant,* and *population* across different genres in the Corpus of Contemporary American English (COCA).
2. To examine the degree of formality of each term.
3. To analyze the collocational tendencies of each term and identify common lexical partners.
4. To explore the semantic preferences in which these terms commonly occur.

1.3 Research Questions

This study seeks to answer the following research questions:

1. What are the frequency patterns of *resident, inhabitant,* and *population* in different genres within COCA?
2. How does the degree of formality differ among *resident, inhabitant,* and *population* across various COCA genres?
3. What are the common collocational patterns associated with each term?
4. What semantic preferences characterize the use of these terms?

1.4 Definitions of Key Terms

Synonymy: The semantic relationship between two or more words that share similar meanings.

Near-Synonyms: Words that have similar core meanings but differ in connotation, collocation, register, or usage context.

Corpus Linguistics: A research method that uses large collections of authentic texts (corpora) to study patterns of language use.

Collocation: The habitual pairing or co-occurrence of words within a language.

Semantic Preference: The tendency of certain words to appear within specific thematic or semantic fields.

Frequency Analysis: A method used in corpus linguistics to measure how often a word or phrase appears within a corpus.

1.5 Scope of the Study

This study is focused on near-synonyms, including *resident*, *inhabitant*, and *population*, analyzing their frequency, collocations, and semantic preferences drawn from COCA evidence. The study only takes into account American English in the analysis because COCA essentially reflects this variety. Even though the study provides in-depth accounts of the usage of the target words, it does not include all possible synonyms for habitation and population.

Thus, the findings cannot be automatically applied to other varieties of English such as British or Australian English as there could be differences in usage across regions. Furthermore, the study is only based on genres that are available in COCA and is not representative of informal or non-standard English commonly found in social media or internet discourse.

1.6 Significance of the Study

This research contributes to the ELT field by making empirical comments on the different shades of meaning between *resident, inhabitant,* and *population*. From its corpus-informed approach, the research moves beyond the confines of dictionary meanings and examines how these terms function in actual contexts. The findings are expected to help both learners and educators. Learners will have enhanced understanding of the near-synonyms, which will enhance lexical skills and enhance communicative accuracy. For educators, the study provides evidence-based teaching guidelines for vocabulary teaching to enable improved pedagogy that addresses common student difficulties. In addition, this research contributes to the field of lexical semantics in general through a careful study of synonymy, highlighting the role of corpus linguistics in distinguishing between close-synonyms. The study also has implications for practical application in curriculum developers, lexicographers, and material developers seeking to create resources that better respond to the complexity of synonymy in English.

CHAPTER 2
REVIEW OF LITERATURE

2.1 Introduction

The study of synonymy is a central study in lexical semantics, particularly in English Language Teaching (ELT), because accurate use of vocabulary is very important in effective communication. In second language acquisition, learners often experience difficulty with in-depth differences between near-synonyms, and a breakdown or misinterpretation results. *Resident*, *inhabitant*, and *population* are words that have similar core meanings but context-dependent applications. This chapter follows the theoretical and empirical history of synonymy, grounding its arguments in linguistic theory and corpus-based studies to elucidate how such near-synonyms may be distinguished. It follows definitions and approaches to synonym, criteria for distinguishing near-synonyms, corpus linguistics and word studies, and lessons from previous research. Through this protracted overview, we lay the groundwork for the ensuing corpus-based analysis, referring to both the importance of contextual differentiation of synonyms and to the robustness of empirical proof in uncovering fine-grained pragmatic and semantic differentiation.

2.2 The Concept of Synonymy in Linguistics

2.2.1 The Definitions of Synonyms

Synonymy has traditionally been defined in linguistics as the relation between similarity or identity of meaning among words. The Oxford Advanced Learner's Dictionary (9th ed., 2015) offers a definition of a synonym as "a word or expression that has the same or nearly the same meaning as another in the same language". This definition highlights that not complete identity of meaning is required; words can be synonyms if they share a kernel of similar meaning. The term synonym itself comes from Greek syn- ("with/together") and -onym ("name"), which implies the idea of words sharing a common name or reference to a thing. Historically, English inherited numerous synonyms because it had borrowed from French, Latin, and other sources

(e.g., the native ask and the French-derived inquire), leading to sets of two words denoting the same sense (Palmer, 1997).

2.2.2 Absolute Synonyms, Strict Synonyms and Loose Synonyms

Notwithstanding this profusion, linguists have long commented that absolute synonymy—wherein two words are identical in all meanings and uses—is very rare (or indeed nonexistent). If two words were perfectly synonymous in use and meaning, linguistic economy would mean that one would eventually fall into disuse or be redefined (Martinet, 1955, quoted in Vicentini 2003; see also Clark 1992) As Cruse (1986) says, two lexical units could be absolute synonyms only "if and only if all their contextual relations were identical"– something which is hardly ever true in natural language. Rather, the majority of synonyms have minor differences in meaning, collocation, register, or grammar and are thus near-synonyms and not perfect substitutes (Lyons, 1995). For instance, take the words begin and commence.

While both verbs are employed in initiating an action, commence is more formal and frequently applied in legal, scholarly, and commercial discourse (i.e., The ceremony will start at noon), while begin is more neutral and frequently encountered in spoken language (i.e., Let's start the meeting). Such subtle variation in use underscores the necessity for investigating synonymy beyond dictionary definitions. Jackson and Amvela (2000) have termed such theoretical examples strict synonyms, i.e., words that can be substituted in any context without any change in denotation, connotation, or stylistic effect . Strict synonyms are not available as a matter of fact since no two words have all the same linguistic features . For example, even if terms denote the same thing, they may vary in register or frequency: groundhog and woodchuck denote the same animal, but it might be used in scientific applications while the other is applied colloquially, or one is more regional.

These instances (typically with technical jargon or dialectic forms) are as near to absolute synonymy as language gets. Ultimately, one of the terms will gain a distinct connotation or field of application so that there is never total overlap. Linguist Eugene Nida metaphorically states that two perfectly substitutable words would be "wasteful" within a language system because they would be repeating the same thing twice. Far more common are near-synonyms, also known as loose synonyms or cognitive synonyms, which share a great deal of meaning but differ in small ways. Near-

synonyms are "words that are extremely close in meaning but cannot be substituted for each other in every situation.". Cruse states that they differ from one another in their peripheral semantic features but are identical in their central semantic features.

Such variations may be with reference to connotation, collocational range, style level, or subtle shades of meaning. For instance, the adjectives beautiful, attractive, lovely, and charming all overlap in expressing favorable assessment of look, but each of them has an insistent nuance of usage and tone (e.g., charming rather implies a charming character, while attractive is more formal). The important point here, though, is that near-synonyms are not fully interchangeable: it is not possible to switch between them freely without danger of altering the delicacy or acceptability of a sentence . A classic example is the set find vs. discover . Both verbs both mean encountering something and therefore share a core denotative meaning. Discover, however, tends to typically mean initially locating something or uncovering something hidden or unrevealed, whereas find could be more neutral (simply encountering something or acquiring an experience). In a specific context like "Marie Curie discovered radium in 1898," the verb discovered cannot be replaced with found without a change in implication. (Marie Curie found radium may mean quite possibly finding it, with the lack of subtlety of an advance on behalf of science.) This illustrates that near-synonyms can have overlapping meanings but divergent entailments or presuppositions.

Similarly, while big and large are typically established as synonyms, they are not always interchangeably substitutable in any collocations: a native speaker employs "What's the big deal?" to convey "Why is it important?", whereas employing "large deal" in the same idiom would be odd . The fixed colloquial meaning of the big deal combination is not available with large deal . Such facts reveal the function of context and conventional use in synonymy. The substitutability test—– whether a word can be exchanged for another without altering acceptability or sense—– is a diagnostic test for synonymy. Near-synonyms will pass this test in at least some contexts, highlighting the fact that their senses are only nearly identical, and not identical. Overall, synonymy exists on a scale: at one end, quite rare examples of (near) total overlap; and at the other, clearly different words; and in the vast majority, near-synonymous sets or pairs with shared meaning but differing in one or more facets of language (Lyons, 1995; Murphy,

2003). Linguists are interested in these fine distinctions, since they need to know them in order to use language accurately and naturally.

2.3 Criteria for Distinguishing Near-Synonyms

Given that most so-called synonyms are actually near-synonyms, it becomes essential to establish criteria for distinguishing them. Researchers have identified a range of linguistic factors that differentiate words with similar meanings (Cruse, 1986; Palmer, 1981; Murphy, 2003). These include connotation, stylistic level, collocational behavior, grammatical patterns, and more.

2.3.1 Connotation and Emotive Tone

Another significant factor in distinguishing near-synonyms is connotation, which refers to the emotional, cultural, or attitudinal associations that a word carries beyond its dictionary definition. Even when two words share a similar denotative meaning, their connotations may evoke different impressions, levels of intensity, or subjective evaluations (Leech, 1981).

For example, the adjectives *cheap* and *inexpensive* both describe something that costs little money, but *cheap* often carries a negative connotation, implying poor quality or lack of durability (*The shoes were cheap, but they fell apart quickly*), whereas *inexpensive* suggests affordability without necessarily implying low quality (*She found an inexpensive hotel that was still very comfortable*).

Similarly, consider the difference between *stubborn* and *determined*. Both adjectives describe someone who does not change their opinion easily, but *stubborn* has a negative connotation, suggesting inflexibility or unwillingness to listen (*He was too stubborn to admit he was wrong*), while *determined* conveys a positive connotation, implying perseverance and resilience (*She was determined to finish the marathon despite the challenges*).

This difference in connotation is also evident in verbs. For instance, *steal* and *borrow* both involve taking something, but *steal* implies illegality and wrongdoing (*He was caught stealing from the store*), whereas *borrow* suggests temporary use with permission (*She borrowed a book from the library*). The choice of one word over another depends not only on meaning but also on the speaker's intent and the situational context.

Below is a list of words that share similar meanings but have distinct connotations:

- Job vs. Career – *Job* refers to employment in a general sense (*She found a job at the café*), while *career* implies long-term professional growth (*He built a successful career in medicine*)
- House vs. Home – *House* denotes a physical structure (*They bought a new house*), while *home* carries emotional significance and warmth (*There's no place like home*)
- Talk vs. Converse – *Talk* is informal and common in daily communication (*Let's talk about it later*), whereas *converse* is more formal and structured (*They conversed about politics at length*)

Understanding these distinctions is crucial for language learners and linguists alike, as incorrect word choice can lead to unintended meanings. Since connotation often reflects societal attitudes, cultural differences may also influence how certain words are perceived in different languages and regions. This aspect of synonymy demonstrates the importance of studying connotational differences within a broader linguistic framework.

2.3.2 Degree of Formality (Register)

Coming to another crucial factor in distinguishing near-synonyms, it is the degree of formality, which refers to the level of politeness, professionalism, or social appropriateness associated with a word. Some words are more appropriate in formal, academic, or professional settings, while others are commonly used in casual, conversational, or colloquial speech (Leech & Svartvik, 2003). Understanding the level of formality of a word helps speakers and writers choose the appropriate expression based on the communicative context.

Formal and informal synonyms often have similar meanings but differ in their degree of politeness, register, and usage context. For example, the verbs *commence* and *begin* both indicate the start of an action, but *commence* is more formal and typically found in official or academic discourse (*The ceremony will commence at noon*), whereas *begin* is neutral and commonly used in everyday conversation (*Let's begin the meeting*).

Similarly, the words *purchase* and *buy* both refer to acquiring something in exchange for money, but *purchase* is more common in legal or commercial contexts (*The company purchased new equipment*), while *buy* is preferred in casual speech (*I bought a new phone yesterday*).

The distinction between formal and informal words is also evident in adjectives. For instance:

- ∉ *Deceased* (formal) vs. *dead* (neutral)
- ∉ *Illness* (formal) vs. *sickness* (informal)
- ∉ *Utilize* (formal) vs. *use* (neutral)

These examples illustrate that while formal and informal synonyms share similar meanings, their appropriateness depends on the context, audience, and communicative intent (Palmer, 1997).

2.3.3 Contextual Appropriateness and Register

Register refers to the level of formality or informality in language use, which can be categorized into different domains, such as:

1. Academic and Legal Register – Uses highly formal vocabulary appropriate for professional, legal, or scholarly settings.

   Example: *The defendant absconded from custody.* (*Absconded* is formal and legalistic, whereas *ran away* is informal.) (O'Keeffe & McCarthy, 2010)

2. Business and Official Communication – Prefers semi-formal expressions that maintain professionalism.

   Example: *The organization seeks to rectify the issue.* (*Rectify* is formal, whereas *fix* is informal.) (Xiao & McEnery, 2006)

3. Everyday Conversation – Uses casual and informal vocabulary appropriate for personal interactions.

   Example: *She looked into the problem.* (*Looked into* is informal, whereas *investigated* is more formal.) (Phoocharoensil, 2010)

A mismatch between formality and context can lead to awkward or inappropriate communication. For instance, using an overly formal term in a casual conversation may sound pretentious or unnatural, while using an informal word in a

professional setting may seem unprofessional or imprecise (Sakulratanacharoen, 2013).

Below is a list of words that vary in their degree of formality:

- Ask (neutral) vs. Inquire (formal)
  – *She asked about the report* vs. *She inquired about the report.* (Cruse, 1986).
- Get (neutral) vs. Obtain (formal)
  – *He got the certificate* vs. *He obtained the certificate.* (Palmer, 1997).
- Help (neutral) vs. Assist (formal)
  – *Can you help me?* vs. *Can you assist me?* (Lyons, 1995).
- Fix (neutral) vs. Repair (formal)
  – *He fixed the car* vs. *He repaired the vehicle.* (Leech, 1981).
- Start (neutral) vs. Commence (formal)
  – *The meeting will start at 10 a.m.* vs. *The meeting will commence at 10 a.m.* (Murphy, 2003).
- End (neutral) vs. Terminate (formal)
  – *The contract ended last month* vs. *The contract was terminated last month* . (Sinclair, 1991).
- Give up (neutral) vs. Relinquish (formal)
  – *He gave up his position* vs. *He relinquished his position.* (O'Keeffe & McCarthy, 2010).
- Tell (neutral) vs. Inform (formal)
  – *She told me about the changes* vs. *She informed me about the changes.* (Xiao & McEnery, 2006).
- Need (neutral) vs. Require (formal)
  – *The project needs more funding* vs. *The project requires additional funding* .(Phoocharoensil, 2010).
- Use (neutral) vs. Utilize (formal)
  – *We use advanced technology* vs. *We utilize advanced technology* .(Sakulratanacharoen, 2013).

These examples highlight the importance of selecting the appropriate degree of formality based on the situation and audience. For second-language learners,

mastering these distinctions is essential for effective communication in diverse contexts (Cruse, 2004).

2.3.4 Vocabulary Differences Across Dialects

Many near-synonyms exhibit dialectal differences, with certain words being preferred in one variety of English over another. For instance, in American English, the word *apartment* is used to describe a self-contained residential unit, whereas in British English, the term *flat* is preferred (Crystal, 2003). Similarly, Americans commonly use *elevator*, while British speakers favor *lift* (Algeo, 2006). These distinctions demonstrate how regional variation can influence synonym selection.

Other examples include:

- Truck (AmE) vs. Lorry (BrE) – *He loaded the truck with furniture* vs. *He loaded the lorry with furniture* (Merriam-Webster Dictionary, 2023).
- Fall (AmE) vs. Autumn (BrE) – *The leaves change color in the fall* vs. *The leaves change color in the autumn* (Oxford Learner's Dictionary, 2023).
- Pants (AmE) vs. Trousers (BrE) – *He bought new pants from the store* vs. *He bought new trousers from the shop* (Collins English Dictionary, 2023).

These differences are not only lexical but can also reflect cultural and historical influences on language development.

2.3.5 Spelling Variations Between British and American English

Spelling variations are another form of dialectal divergence that affects synonymy. British and American English follow different conventions due to historical developments and standardization processes led by figures such as Noah Webster in the U.S. and Samuel Johnson in the U.K. (Crystal, 2012). Examples of spelling differences that may affect synonym usage include:

- Colour (BrE) vs. Color (AmE) – *The new design features vibrant colours* vs. *The new design features vibrant colors* (Merriam-Webster Dictionary, 2023).
- Organise (BrE) vs. Organize (AmE) – *She helped organise the event* vs. *She helped organize the event* (Cambridge Dictionary, 2023).
- Defence (BrE) vs. Defense (AmE) – *The team's defence was strong* vs. *The team's defense was strong* (Collins English Dictionary, 2023).

Although these spelling differences do not significantly impact meaning, they can influence word choice and formal writing preferences in different regions.

2.3.6 Collocation

Collocation is a fundamental aspect of vocabulary acquisition and lexical semantics. It refers to the predictable association of words that frequently appear together in natural language use (McCarthy, 2006). Unlike simple word combinations, collocations are not always grammatically required, but they form strong linguistic patterns that native speakers recognize intuitively.

For example, while both *strong* and *powerful* convey a sense of intensity, they collocate differently:

- *Strong coffee* (common) vs. *Powerful coffee* (uncommon)
- *Powerful engine* (common) vs. *Strong engine* (less common)

Similarly, synonyms such as *make* and *do* are distinguished by their typical collocations:

- *Make a decision* (common) vs. *Do a decision* (incorrect)
- *Do homework* (common) vs. *Make homework* (incorrect)

Such distinctions illustrate that even words with similar meanings cannot be freely interchanged without considering their collocational constraints (Nation, 2001).

2.3.6.1 Types of Collocations Collocations can be classified into several types based on the relationship between words. Some of the most commonly studied collocation types include:

1. Adjective + Noun – *deep regret, heavy rain, strong argument* (Biber et al., 1999)
2. Verb + Noun – *commit a crime, make progress, take responsibility* (Sinclair, 1991)
3. Noun + Noun – *brain drain, job market, stock exchange* (McCarthy, 2006)
4. Verb + Adverb – *work hard, argue strongly, contribute significantly* (O'Keeffe et al., 2007)
5. Adverb + Adjective – *highly successful, deeply concerned, widely recognized* (Biber et al., 1999)

Collocational restrictions exist due to both semantic compatibility and language convention. While non-native speakers might assume that certain words can be substituted freely, native speakers instinctively know that only specific combinations sound natural (Sinclair, 1991).

2.3.6.2 Collocational Differences Between Synonyms Collocational behavior is often a key factor in distinguishing synonyms. Words with similar meanings may exhibit different preferences in terms of the words they co-occur with (McEnery & Hardie, 2012). Below are some examples of synonym pairs and their preferred collocations:

- Big vs. Large
    - *Big mistake* (common) vs. *Large mistake* (less common)
    - *Large quantity* (common) vs. *Big quantity* (uncommon)

(Oxford Collocations Dictionary, 2023)

- Begin vs. Start
    - *Begin a project* (common) vs. *Start a project* (also common)
    - *Start a car* (common) vs. *Begin a car* (incorrect)

(Cambridge Dictionary, 2023)

- Speak vs. Talk
    - *Speak a language* (common) vs. *Talk a language* (incorrect)
    - *Talk nonsense* (common) vs. *Speak nonsense* (less common)

(Merriam-Webster Dictionary, 2023)

2.3.7 Grammatical Patterns

Grammatical patterns refer to the structural rules that govern how words are used within sentences. These patterns include variations in verb valency, transitivity, prepositional usage, and clause structures (Biber et al., 1999). Understanding these patterns is essential for language learners, as misapplying a grammatical structure can result in ungrammatical or awkward sentences.

For example, consider the near-synonyms *suggest* and *recommend*. While both verbs indicate giving advice or proposing something, they differ in their grammatical constructions:

- *She suggested going to the beach.* (Correct)
- *She recommended going to the beach.* (Correct)
- *She suggested to go to the beach.* (Incorrect)
- *She recommended to go to the beach.* (Correct)

(Oxford Learner's Dictionary, 2023)

These examples illustrate that *suggest* requires a gerund (-ing form), whereas *recommend* can take either a gerund or an infinitive (to + verb) in certain contexts (Cambridge Dictionary, 2023).

2.3.7.1 Transitivity and Verb Agreement Structures One major grammatical distinction between synonyms lies in transitivity, which refers to whether a verb requires a direct object (transitive) or can stand alone without one (intransitive) (Levin, 1993). Consider the verbs *rise* and *raise*:

- *The sun rises at 6 a.m.* (Intransitive: no object)
- *She raised her hand to ask a question.* (Transitive: requires an object)

(Merriam-Webster Dictionary, 2023)

Similarly, verbs like *explain* and *describe* require different grammatical structures:

- *She explained the rules to us.* (Verb + direct object + indirect object)
- *She described the situation.* (Verb + direct object only)

(Collins English Dictionary, 2023)

Understanding such transitivity constraints is crucial in distinguishing synonyms with different valency properties.

2.3.7.2 Prepositional Usage with Synonyms Many synonyms differ in their prepositional requirements, which dictate which prepositions can accompany them in a sentence. This is particularly common in verb-preposition constructions. Consider the verbs *rely* and *depend*:

- *She relies on her team for support.* (Correct)
- *She depends on her team for support.* (Correct)
- *She relies of her team for support.* (Incorrect)

(Oxford Collocations Dictionary, 2023)

Prepositional variation is also seen in adjectives:

- *Good at math* vs. *Interested in math*
- *Familiar with the subject* vs. *Aware of the issue*

(Cambridge Dictionary, 2023)

These restrictions highlight how preposition choice affects the grammatical acceptability of synonyms, reinforcing the idea that synonymous words are not always interchangeable.

2.3.7.3 Passive vs. Active Voice Constraints Another syntactic factor distinguishing synonyms is their compatibility with passive constructions. Some verbs are more naturally used in passive voice, while others are typically found in active constructions. Consider the verbs *complete* and *finish*:

- *The project was completed on time.* (Common passive usage)
- *The project was finished on time.* (Less common passive usage)

(Merriam-Webster Dictionary, 2023)

Similarly, verbs like *believe* and *think* exhibit different passive voice tendencies:

- *It is believed that he left early.* (Common)
- *It is thought that he left early.* (Common)
- *It is considered that he left early.* (Less common)

(Oxford English Dictionary, 2023)

These distinctions reflect verb-specific preferences for active or passive voice constructions, contributing to grammatical constraints on synonym usage.

2.4 Corpus Linguistics and Word Studies

Both dictionary meanings and the more traditional introspective methods, while useful, are always bound to be too narrow in their attempts to depict the whole range of differences between near-synonyms. Corpus linguistics is where this method truly shines. Corpus linguistics involves studying enormous collections of real texts (corpora) to observe how words are actually used in real usage. By studying natural use on a large scale, researchers are able to identify patterns of frequency, collocation, and syntactic behavior that may go unnoticed or may be impossible to estimate using intuition alone. As Lindquist (2009) highlights, corpus-based methodology gives us an objective empirical footing for the analysis of language patterns with which to transcend anecdotal evidence or prescriptive rules. In synonymy, evidence from the corpus may reveal systematic variation in the distribution of near-synonyms across contexts and consequently their pernicious variations meaning and use. Flowerdew (2012) notes the teaching potential of such analysis on the basis that Data-Driven Learning (DDL) approaches—— in which learners explore corpus examples—— – can largely facilitate awareness of word use. By using a balanced corpus like the Corpus of

Contemporary American English (COCA) or the British National Corpus (BNC), one can quantify frequencies of resident, inhabitant, and population in genres from fiction, newspaper articles, academic writing, to spoken speech.

This type of analysis might show, for example, that inhabitant is far more common in academic and historical language (where talk about settlement and geography dominates), but population might be more common in population studies and sociology, and resident might be found quite generally across genres but especially in legal, housing, or town planning discourse. Frequency alone cannot explain why differences do exist, but it does provide a quantitative profile of use, any theory of synonymy having to account for. It also helps fix on which word is more general or distinctive in typical use (e.g., if one word is significantly less frequent overall, it will perhaps be more limited in meaning or context). One of the largest advantages of corpus analysis is that frequency analysis can be performed. That involves quantifying how often each synonym appears in terms of genre or registers. Frequency counts may give initial hints: e.g., if inhabitant is far more frequent in academic writing than resident, that suggests a register or genre bias. In the same way, if resident is radically more frequent in everyday conversation than inhabitant, that suggests variation in formality or situation of use.

2.4.1 Concordance

Concordancing is another central corpus technique for synonym study aside from frequency. A concordance is a list of citations for a search term with some context (typically shown as KWIC – Key Word in Context – lines). By examining concordance lines for inhabitant, resident, and population, one can at once identify characteristic contexts and collocations for each noun. For example, a KWIC perspective would show how many instances of population preceded by adjectives like growth, density, increase, rural, confirming our initial hunch about collocations. It would perhaps show resident most frequently followed by adjectives like permanent, local, or senior, and inhabitant occurring close to words like original, indigenous, native. Concordancing enables patterns to be identified: one can see that inhabitant scarcely co-occurs with apartment in an extended corpus, but resident does so frequently. The results agree with the collocational discrepancies revealed in 2.3. In addition, concordance lines can indicate semantic prosody—the good or bad flavor a word carries based on its usual

contexts (Louw, 1993). For instance, if population typically appears in situations describing overpopulation, decline, or crisis, it suggests a quite negative or neutral prosody. Near-synonyms might show a prosody difference: one classic example outside of our trinity is slim vs. skinny, where slim positively tends and skinny normally negatively. Concordancing corpus lines for resident/inhabitant/population may uncover differences in tone—though all three are about aggregations of human beings, resident might be used more often in city administration writing, inhabitant in geography descriptions of place, and population in statistical or ecological argumentation. Concordancing thus provides a qualitative glimpse into use as a supplement to quantitative measures.

2.4.2 Collocation

Another robust corpus method is statistical measure-based collocational analysis. Collocation programs can determine the strength of association between the target synonyms and other words, usually measured by metrics like Mutual Information (MI) or t-score.

Corpus analysis can report the strongest collocates of *resident, inhabitant,* and *population* and throw light on their unique phraseologies. For example, it can be discovered that the strongest collocates for resident are apartment, permit, local, home, while for inhabitant they are indigenous, native, early, and for population they are growth, density, census. These discoveries clearly demonstrate how each noun "draws" different linguistic environments. Petcharat (2016), a corpus study of the near-synonyms appropriate, proper, and suitable, showed that even very close dictionary definitions for words do have different collocational profiles showing their functional distinctions. In the same way, corpus-based study of *resident*/*inhabitant*/*population* can elicit patterns not visible from definitions. For instance, *resident* may appear more in law and administrative writing (*residents* of the building must comply with rules), while *inhabitant* appears more in historical or environmental writing (*inhabitants* of the ancient city). It might also make differences in grammatical voice apparent: possibly *population* turns up most frequently in passive structures in formal writing (the population was surveyed), while *resident* appears more often in active structures (residents complained). These refinements justify the use of each noun.

2.4.3 Semantic Preference

A notable point is that corpus research enables researchers to test semantic preference systematically. By tagging or marking the contexts where synonyms appear, it is possible to identify common thematic relations. As observed, *resident* may appear in corpora mostly within legal, administrative, or housing contexts, inhabitant in geographical or historical ones, and population in demographic or sociological writing. Quantifying these tendencies (e.g., population X% of the time in urban planning texts, inhabitant Y% of the time in natural environment texts) confirms our understanding of each word's pragmatic niche.

This is what was achieved in studies such as Krawczak (2014), which examined adjectives of shame (ashamed, embarrassed, humiliated) in two corpora and was able to describe each adjective's preferred contexts (personal guilt vs. social embarrassment vs. public humiliation).

Such findings show that near-synonyms specialize in different shades of meaning within a broad semantic field. In short, corpus linguistics provides frequency evidence (frequency, collocate statistics, distribution by genre) and evidence of use (concordance line analysis of use) to distinguish near-synonyms that introspection or dictionaries may fail to do. Such a data-driven approach has received widespread adoption in the last couple of decades for lexicographical research. It accords with the principle of distributional semantics that "a word is defined by the company it keeps," i.e., that looking at what words are in proximity to a given word tells us much about its meaning and usage. By using corpora like COCA for American English (or the BNC for British English, to name but one), scholars can look at resident, inhabitant, and population in the context of millions of words of text. Patterns which have been extracted from corpora will inform the analysis in Chapter 3, but even at the literature review stage, we see that some corpus-based research has highlighted previously undetailed differences between near-synonyms. For instance, Divjak and Gries (2006) used corpus data and statistical analysis to examine Russian near-synonymous verbs, demonstrating that even closely related syntactic forms exhibit distinct usage patterns when analyzed across thousands of examples.

In the same way, this research calls on corpus evidence to capitalize on the subtleties of *resident*, *inhabitant*, and *population*, going beyond static definitions to dynamic use. As we shall demonstrate, it is this approach that offers high payoff to

language teaching and learning, since it highlights the actual patterns students must learn.

2.5 Previous Studies

Several corpus-based studies have examined synonymy by examining how words of the same sense are used in actual contexts. The section highlights significant studies that have examined synonyms in great detail, their methodologies, findings, and implications to the teaching of English.

One notable study by Petcharat (2016) investigated the synonyms *appropriate*, *proper*, and *suitable* using a corpus-based approach. The study focused on these three adjectives, which are closely related in meaning and all convey the idea of correctness or suitability. Petcharat utilized several learner's dictionaries and the Corpus of Contemporary American English (COCA) to analyze their meanings, levels of formality, collocations, and grammatical patterns. The findings revealed clear differences in spite of the common sense of "acceptable or fitting:": appropriate had been more common in formal and scholarly settings, predisposing to imply suitableness to a purpose or situation (e.g., appropriate measures, appropriate behavior). Correct carried over connotations of old-fashioned formality or correctness and was likely to happen in social norms or normed behavior situations (e.g., correct manners, the correct authorities)—it also possessed a BrE vs AmE use difference in a few situations (e.g., correct as an intensifier in BrE). Appropriate was frequently used in situations of conceptual appropriateness for some end (appropriate action, appropriate response), and was standard when speaking about whether or not something was right by some criteria. Collocationally, they both had favorites: appropriate with words like action, response; proper with name, place, respect (and in phrases like properly etc.); suitable with for phrases (suitable for X). Petcharat's corpus analysis uncovered distinctions not immediately apparent from dictionary definitions in isolation, e.g., the dictionaries can all say "appropriate = suitable = proper (meaning: right/correct for the context)," but the corpus showed each word's behavior pattern was distinct. This is the crucial point to make for our research: although fundamental senses may overlap, usage can vary quite considerably.

Petcharat's research demonstrates that combining dictionary and corpus data provides a more comprehensive understanding of language use. The study's implications for English Language Teaching (ELT) suggest that instructors should teach these adjectives alongside their typical collocational patterns (e.g., *appropriate for/to, suitable for*, but not *proper for* in the same way) and the contextual limitations of their usage. This research is an exemplar, methodologically, of how to proceed *resident*/*inhabitant*/*population*, which we anticipate having similarly nuanced distributions.

Another study by Phoocharoensil (2010) experimented with five request-relation verbs: *ask, beg, plead, request,* and *appeal* using both corpus evidence as well as dictionary definitions. In the research, it was found that while all five are all about requesting, they differ considerably in situation and tone. *Ask* was found to be the common, general word—usable in most contexts for simply asking for something or getting something done. *Beg* and *plead* were more emotional and urgent, with *beg* having connotations of desperation and humility (tending to be found collocating with for mercy, for help etc.), and *plead* tending to be used in law or drama (and generally used with someone or for something in desperation). *Request* was the most formal and polite, often used in writing or polite speech (e.g., *requesting* permission, *request* that a person do something), and typically used for medium to high formality contexts (business letters, announcements). *Appeal* (as a verb sense request) was likely to be an urgent request either publicly or officially, and was often used in a legal or fund-raising situation (e.g., *appeal* to the public for help, *appeal* for restraint), being of a persuasive nature. Syntactically, *appeal* and *request* are likely to be clausal complements (appeal to someone to do something, request someone to do something), *beg* and *plead* may have direct objects or infinitival clauses (plead someone's pardon, beg someone to do something), etc. The corpus data confirmed these differences: *request* was observed mainly in formal writing, *beg* and *plead* occurred most in literature dialogue (mirrored the spoken plea), ask was found everywhere in all contexts. An interesting finding was the semantic prosody and collocation difference: *plead* co-occurred with guilty (as in plead guilty, a legal expression), which differentiates it from the others. *Appeal* occurs together with court, verdict in juridical words and also under urgency, ruling etc. These verbs also reflected differences in grammar, – e.g., one *appeals* to someone for

something, whereas one *begs* someone for something or *begs* for something directly. Phoocharoensil's study indicated that even from a restricted semantic domain (asking for something), each synonym has its own unique profile. For our purposes, this reinforces the importance of considering formal versus informal usage (e.g., *achieve* vs. *get,* or *fulfill* vs. *carry out*), as well as the distinction between emotive and neutral tones. While tone may be less of a concern for relatively neutral or positive nouns like *resident, inhabitant,* and *population,* it becomes more relevant in other categories. The findings also highlight the need to pay attention to argument structure, such as verb patterns.

Moreover, Chung applied comparative analysis of the near-synonyms *create* and *produce* in accordance with the Brown Corpus (American English, 1960s) and Frown Corpus (American English, 1990s). The objective was to see how the "bring something into existence" synonyms differ in semantic and grammatical usage. The findings revealed a clear contextual divergence: *create* tended to occur more in abstract or artistic pursuits, whereas *produce* occurred more in concrete products or production environments. To illustrate, one *creates* art work, stories, conflict, atmosphere, etc., but one *produces* goods, evidence, results. The corpora indicated *create* tend to occur with nouns like problem, jobs, image, impression (non-concrete objects), and *produce* with nouns like car, film, sound, report. There was also a contrast in number and countability of the objects: *create* favored singular, usually abstract nouns (*create* an opportunity), whereas *produce* had mass or plural nouns, usually in factory environments (*produce* cars, produce evidence). Grammatically, *produce* had more passive use in technical writing (e.g., 10 mL of gas was produced), while *create* was used less so. Chung further added *create* implies bringing into existence something new/original (hence usage in creative or inventive contexts), while *produce* is simply producing something, not necessarily new (simply making it present or available). This distinction is subtle but important, and learners often mix them up by using produce in creative contexts or vice versa. The study's implication is that paying attention to collocational range (abstract vs concrete) and domain (art vs manufacturing) helps differentiate synonyms.

Cai's study examined a set of seven adjectives, all generally conveying the meaning of "very good," using the Corpus of Contemporary American English (COCA) to analyze their frequency, collocations, and nuanced meanings. Each of these

evaluative adjectives carried its own connotation. *Great* emerged as the most frequently used term and functioned as a broadly positive, highly context-flexible adjective—appearing in phrases ranging from *"a great time"* to *"a great man."* *Fabulous* and *fantastic*, on the other hand, implied over-enthusiasm or hyperbole, and tended to be found in informal spoken contexts. *Awesome* was found to be informal and especially common in spoken American English (at least in the 2000s, as a stand-alone exclamation "Awesome!"), retaining its etymological meaning of "causing awe" but in everyday use merely "very good/cool.". *Excellent* was largely used in more formal or written contexts, or when assessing performance (e.g., excellent performance, excellent condition), corresponding to the academic or professional environment. *Great* historically meant to cause terror, but now "extremely good," and Cai felt it was used less often by the 2000s except in some fixed expressions (and some older generations). *Wonderful* was reserved for more heartfelt or true praise, usually to describe personal experience or that which literally give wonder/happiness (e.g., wonderful surprise, wonderful feeling). Pragmatic and stylistic distinctions were emphasized in this study, although all were synonyms. Students often ask about the difference because dictionaries might list all of them as "very good (informal/formal)" but not frequency of use or tone. Frequency counts by Cai showed that *great* was most used, *awesome* rising in usage, *excellent*/*wonderful* steady, *fabulous*/*fantastic* being used but somewhat less, etc. The point of interest to notice for our purposes is the manner in which synonyms could differ in level of enthusiasm, formality, and genre.

Sakulratanacharoen (2013) studied the synonyms *remark*, *comment*, and *observe* in journalistic contexts. Corpus analysis took into account adverbial modifiers and ordering structures of reported speech. *Remark* and *comment* were also discovered to be used widely in making statements in news reporting, often replacing said but with minor variations: *remark* can imply a casual/incidental speaking, and was generally followed by adverbs like casually, wryly ("remarked wryly"), while *comment* is a little more formal/neuter, often no adverb or with simple ones like publicly, further ("commented further that."). Used as a reporting verb, *observe* has a more formal and slightly antiquated/learning-sounding tone and generally implies a reflective or analytical utterance ("observed that the trend was alarming"), and it is sparingly used in serious reporting or analysis.

This study identified how even synonyms that approximately translate to "said" are differently weighted: *remark* is lightest, *observe* is heaviest, *comment* in the middle. The examination of where they go in sentences (e.g., he remarked. vs. he noted that .) revealed some patterns of grammar: *observe* with a that-clause, *comment* with quotes (direct or indirect), etc. The implication for our work is that synonyms tend to have varying frequency of occurrence in varying styles (news versus literature) and may also collocate with varying reporting adverbs or have different syntax. While *resident*/*inhabitant*/*population* are not reporting verbs, this paper is just another demonstration of fine differences and the importance of context, and more importantly genre. In addition, Sakulratanacharoen also pointed out that making learners aware of such differences can be useful to writers who wish to write more native-like text (e.g., students of journalism deciding on remarked vs commented).

Thamratana (2013) investigated five synonyms expressing the concept of reduction: *reduce*, *decrease*, *diminish*, *dwindle*, and *decline*. By analyzing their frequency, grammatical patterns, and collocations using corpus data, the study revealed important distinctions. *Reduce* and *decrease* are commonly used in scientific or technical contexts, often in the passive voice (e.g., *was reduced*, *has decreased*), and convey a neutral, process-oriented tone. In contrast, *diminish* and *dwindle* tend to be more figurative or expressive. *Diminish* is frequently used with abstract nouns such as *hope* or *authority*, often implying a gradual fading. *Dwindle* typically occurs with plural or countable nouns and suggests a steady reduction, often with a tone of concern (e.g., *money dwindled, the population is dwindling*).

Decline is notable for its dual function as both a transitive and intransitive verb. It can indicate polite refusal or the loss of something. In the context of loss, it is primarily intransitive (e.g., *attendance declined*) and appears frequently in formal writing and reports, especially in collocations such as *sharp decline* or *steady decline* (as a noun). The study highlighted that *decline* stands out due to its frequent use as both a noun and a verb, often signaling a downward trend in data. Meanwhile, *reduce* is typically causative and transitive (e.g., *reduce expenses*), whereas *decrease* is more often intransitive (e.g., *expenses decreased*).

These grammatical differences (transitive vs intransitive tendencies) and contextual preferences (technical vs. narrative) again mirror what we've discussed for

our target verbs. Krawczak (2014) compared three adjectives describing shame feelings both on the basis of COCA and the BNC (British National Corpus) for documenting transatlantic usage The results portrayed that the near-synonyms are matched to different intensities and social circumstances of shame. *Ashamed* is generally used when the speaker is guilty or responsible for badness (a personal moral fault: I'm *ashamed* of what I did), tending to collocate with of and a verb or noun (*ashamed* of oneself, ashamed of cheating). *Embarrassed* is social embarrassment or mild shame—typically something which makes one blush or go silly but not necessarily bad (e.g., *embarrassed* about making a public error). It occurs with words like a little, very, *embarrassed* about/at. Stronger than *humiliated* is, which gives a feeling of public shame or debasement of one's dignity, often brought on by others (He felt *humiliated* by the public humiliation). It's used in situations of serious abasement.

The corpora confirmed that *ashamed* is used most frequently in first person (people admitting shame), *embarrassed* in ordinary circumstances and most frequently with bodily expressions signs (blushing, looking *embarrassed*), *humiliated* in cases of abuse or failure witnessed by others (collocations with feel, completely *humiliated*, *humiliated* by). It indicates the importance of emotional connotation and extent in identifying synonyms. It also used two corpora to ensure that the results were not locale-specific; in the same vein, one could check *resident*/*inhabitant*/*population* in American and British corpora for comparison .

Krawczak's use of both the COCA and BNC corpora results in a comprehensive approach that can serve as a model for similar studies. Aroonmanakun (2015) analyzed the adjectives *quick* and *fast* to explore their semantic distinctions and collocational patterns. Although both words convey the idea of "moving or happening at high speed," the study aimed to identify subtle differences in usage and semantic preference. The findings revealed that, while both adjectives can describe speed, they tend to collocate with different nouns and appear in distinct contexts.

Quick is more commonly associated with actions and processes—often implying something done in a short time or with minimal delay (e.g., *a quick look, quick response, quick thinking*). In contrast, *fast* is typically used with physical objects or entities in motion (e.g., *fast car, fast runner, fast pace*). Interestingly, while phrases like *fast food* refer more to the speed of service than the food itself, they are idiomatically

fixed expressions. The phrase *fast snack* might exist but carries a slightly different nuance, lacking the compound and cultural associations of *fast food* (e.g., convenience, low nutritional value). Additionally, *quick* tends to describe individual, often brief actions (e.g., *a quick shower*), whereas *fast* refers more to sustained speed or rate over time (e.g., *a fast car* capable of maintaining high speed).

These tendencies of collocation distinguish *quick* and *fast* in the majority of cases (a native speaker feels a hair's breadth: *fast* response isn't idiomatic, you'd say *quick* response; in contrast, *quick* runner sounds unnatural, you say *fast* runner). The corpus data found strong opposition: *fast* with transport, animals, sport (*fast* train, *fast* horse, *fast* ball), *quick* with activities and subjective judgments (*quick* solution, *quick* conversation, etc.). The lesson being that while synonyms are substitutable (both can be utilized to describe some things like *fast*/*quick* learner are both viable with minor alteration), their most preferred areas of application are different. Much as *resident* and *inhabitant* can both be employed in some contexts, essentially the same sense, but one might be more preferred in a specific construction).

Aroonmanakun's work confirms the role of collocational conventions: they do not necessarily make logical sense (there's no reason why we cannot employ fast answer, but it's adopted as quick answer), and so learners must learn to do so enough as a matter of usage patterns. The review of previous literature in synonymy reveals several major findings and approaches that have enriched the area. Synonym differentiation studies have typically been focused on such issues as semantic difference, register change, patterns of collocation, and syntactic behavior. Corpus linguistics has provided evidence that even near-synonyms exhibit fine-grained yet important differences in usage, which can be researched systematically using corpus linguistics methods. Through comparison of word frequency, collocational patterns, and grammatical preferences, previous studies have successfully demonstrated how near-synonyms differ in real discourse.

A typical thread in previous research is the role of context conditions in deciding among synonyms. Different studies have highlighted that synonyms never must always be spoken since meaning close words may succumb to register-specific boundaries, collocational constraint, and discourse-level influences. Parallel to this, corpus-based research has provided empirical evidence highlighting distinctions between near-

synonyms, demonstrating that word choice is not solely governed by dictionary definitions. While these studies have greatly enhanced our understanding of synonymy, notable gaps remain in the literature—particularly regarding large-scale, corpus-based investigations of less commonly studied synonym pairs. Much of the existing research has focused on high-frequency or prominent synonyms, leaving lower-frequency items under-explored and lacking in robust empirical analysis. Additionally, the majority of studies are based on corpora drawn from limited regional varieties of English, indicating a need for broader, cross-dialectal research to capture more diverse patterns of usage.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

The chapter explains the research design, data collection practices, and analysis practices which are to be used to investigate differences between the near-synonyms resident, inhabitant, and population. Adopting the corpus-based approach, this study aims to provide empirical evidence into the frequency, collocational preferences, grammatical preferences, and semantic preferences for these words. The chapter also discusses the selection of the data sources, the tools to be used for the analysis, and the rationale for the chosen methodology. Utilizing widely established corpus linguistic techniques, the research in this work will try to bring to light fine shades of meaning and usage differences which may not be evident in traditional linguistic analysis.

3.2 Selected Words and Justification

The decision to take *resident*, *inhabitant*, and *population* as the topic of this study is based on their frequent use in legal, demographic, and sociolinguistic contexts, and on the subtle semantic and syntactic distinctions between them.

They are commonly encountered in discussion of geographical location, census returns, and collective identity, and are therefore ideally suited to a detailed corpus-based analysis. Though they are almost synonymous, their usage frequency is controlled by collocational patterns and register, causing learner confusion and non-native speaker misinterpretation. The differences between these words extend beyond semantics to include grammatical properties, collocational constraints, and biases linked to registers. By studying these words with a corpus orientation, this study will endeavor to provide an all-encompassing and empirically credible investigation of their usage patterns, thus supporting the broader field of lexical semantics and synonymy research.

3.3 Data Collection

3.3.1 Research Instruments

To conduct this study, one primary corpus tool will be used: Corpus of Contemporary American English (COCA). This tool enables data extraction, frequency lists, collocational analysis, and semantic preferences.

3.3.1.1 Corpus of Contemporary American English (COCA) The Corpus of Contemporary American English or COCA was used as the primary source of data in this study. The corpus grants access to a huge range of texts representing a variety of registers and is thus best for the study of the distribution of *inhabitant, resident*, and *population* across genres. COCA's in-built search capabilities are utilized in extracting word frequencies, collocations, and grammatical patterns. The corpus permits register-based study in that the corpus will be used to determine where each target noun occurs most commonly and with what.

3.3.2 Corpus Selection

The Corpus of Contemporary American English (COCA) has been chosen due to its size, variety, and dependability. COCA consists of over one billion words and is among the biggest balanced corpora of American English. It contains texts from spoken language, fiction, popular magazines, newspapers, and scholarly writing so that register-based comparison of noun use can be conducted.

Apart from this, COCA has a wide chronological scope with texts spanning from 1990 to date, making it possible to study their long history of usage change. Due to its high availability and reliability, it is an essential tool for linguistic research with accurate frequency counts and collocational patterns ready for empirical analysis.

3.3.3 Data Extraction

In order to extract useful data from COCA, there was a systematic search approach. Each of the three target nouns (inhabitant, resident, and population) was queried on COCA's lemma search tool, which will return all the inflected forms such as inhabitants, residents, and populations.

Each noun's frequency was then recorded across COCA's five main registers: academic, newspaper, magazine, fiction, and spoken.

A collocate was is then conducted using COCA's collocate feature with a span of ±4 words to identify frequent adjective and verb collocates.

In addition, syntactic pattern analysis was also conducted with the help of COCA's part-of-speech filter to examine how these nouns are used syntactically, e.g., how they are complemented by adjectives, preposition phrase patterns, and how they behave in compound noun constructions. Lastly, a semantic preference analysis was conducted through manually examining concordance lines to discover which of these nouns have a tendency to occur primarily in certain semantic contexts.

3.3.4 Data Analysis

3.3.4.1 Frequency Analysis Lexical frequency analysis formed an important component of this study as it will uncover how frequently each of the nouns occurs in the different registers. Frequency analysis included absolute frequency, which recorded raw frequency with which every noun appears in COCA, and normalized frequency, which will calculate frequency per million words in order to allow for informative cross-register comparison (oral speech vs. academic writing) that will also help determine usage trends and potential contextual preferences. The study also examined register-based distribution to determine whether the nouns exhibit register-specific patterns—for example, *inhabitant* appearing more frequently in formal writing, while *resident* is more commonly used in spoken language.

3.3.4.2 Collocational Analysis Collocational analysis also formed a major part of the study in order to determine usual combinations of words and lexical partners for each word. Based on statistical indicators such as Mutual Information (MI) scores, the study measured the extent of association between the target words and the collocates. High MI scores indicated high, statistically significant collocational associations. For example, the analysis specifically focused on identifying frequent noun-adjective collocations such as permanent resident and native inhabitant, and verb-noun collocations to denote activities related to these nouns, such as increase in population or residents. Prepositional collocations were also examined to see which prepositions tend to occur with these nouns, such as resident of, inhabitant in, and population of. Through these collocations, the study establishes what each of the nouns habitually co-occurs with and, in the process, clarify their semantic and syntactic usage.

3.3.4.3 Semantic Preference Analysis Another important analytical dimension is semantic preference analysis, which explores the conceptual categories that nouns tend to collocate with based on typical usage patterns. This study investigates whether each noun appears within specific semantic environments. For instance, *inhabitant* is often associated with geographical or environmental contexts (e.g., *inhabitants of a place, inhabitants of a forest*), whereas *population* more frequently occurs in statistical or policy-related contexts (e.g., *aging population, urban population*). These distinctions enable the study to offer a more fine-grained account of synonym differentiation.

CHAPTER 4
FINDINGS

4.1 Frequency and Formality

Table 1

Overall Frequency and Distribution of Resident, Inhabitant and Population Across the Eight Genres From COCA

Genre	Population		Genre	Resident		Genre	Inhabitant		Total
	Frequency	Per Million		Frequency	Per Million		Frequency	Per Million	
Academics	40300	336.42	**News**	6269	54.49	**Academics**	135	1.13	
Web	16879	135.84	**Web**	2587	20.82	**Web**	112	0.90	
Magazines	14811	117.46	**Magazines**	2570	20.38	**Fiction**	94	0.79	
Blog	14604	113.35	**Academics**	2397	20.01	**Magazines**	67	0.53	
News	13453	110.50	**Blog**	2515	19.55	**Blog**	54	0.42	
Spoken	7679	60.88	**Spoken**	2244	17.79	**News**	30	0.25	
Fiction	2093	17.69	**Fiction**	1098	9.28	**TV**	22	0.17	
TV	1630	12.73	**TV**	1036	8.09	**Spoken**	8	0.06	
Total	**111449**		**Total**	**20716**		**Total**	**522**		**132687**

The table represents the total frequency and trends of the words *resident*, *population*, and *inhabitant* across eight COCA genres. These include blogs, web pages, television, speech, fiction, magazines, news, and scholarly writing. The figures indicate the frequency with which each of the words is used, along with how formal and what type of situations the words are employed in.

In terms of overall frequency, the most utilized of the three by far is *population* at 111,449 occurrences, and then *resident* at 20,716, and lastly *inhabitant* at a mere 522. This suggests that *population* is employed more extensively, likely because it is commonly employed in demographic, statistical, and policy discussions. *Resident* is of

medium frequency and will likely be favored in civic issue situations, residence statute, and domestic environments. In comparison, *inhabitant* appears most infrequently, which implies that it is more specialized and maybe more formal. At the genre-specific level, *population* is more frequent in academic writing at 40,300 occurrences with a high frequency rate per million of 336.42, signifying a strong preference to formal research-based writing. Most frequently, it appears in web magazine, blog, and news genres, all which are prone to reports, data, or discussion of social issues. On the other hand, *resident* is predominantly used in news (6,269 times) and also repeated across other genres such as web pages, magazines, blogs, and academia, meaning that it is widespread in both formal and semi-formal contexts.

Resident, on the other hand, is used more frequently in academic prose (135 times), followed by web pages and fiction. This division of usage implies that *inhabitant* primarily occurs in more formal or descriptive contexts, such as historical or scientific language, compared to more informal or conversational use. It is used infrequently in spoken context, at eight uses, indicative of its comparatively minor usage in informal context.

While *resident*, *inhabitant*, and *population* are close synonyms, their genre distribution and frequency suggest they are used for different communicative purposes. *Population* is the most general and statistical, dominant in data-based and academic genres. *Resident* is common in news and semi-formal texts, often linked to legal or geographical situations. *Inhabitant* is the most formal and least frequent, occurring chiefly in academic or literary language.

4.2 Collocatational Pattern

Table 2

Adjective Collocations of Resident, Inhabitant, and Population From COCA

Rank	Resident			Inhabitant			Population		
	Adjective collocate	**Frequency**	**MI-Scores**	**Adjective collocate**	**Frequency**	**MI-Scores**	**Adjective collocate**	**Frequency**	**MI-Scores**
1	local	2034	4.21	local	151	3.57	large	4427	3.52
2	longtime	784	6.35	original	146	4.63	general	4024	4.65
3	permanent	779	5.65	native	95	4.83	total	2033	4.35
4	legal	612	3.38	indigenous	64	6.06	entire	1758	3.72
5	rural	357	4.28	rural	48	4.51	growing	1397	4.60
6	urban	294	3.55	ancient	47	4.12	civilian	1155	5.91
7	lifelong	286	5.89	remaining	20	3.33	diverse	1102	5.32
8	nearby	273	4.34	sole	20	4.55	native	837	3.85
9	elderly	267	4.62	fellow	18	3.04	urban	818	4.04
10	low-income	258	5.86	peaceful	14	3.71	rural	801	4.46
11	coastal	121	4.08	aboriginal	14	6.21	aging	716	6.27
12	lawful	110	5.97	surrounding	10	3.73	vulnerable	661	4.69

13	year-round	110	5.59	prehistoric	9	5.81	indigenous	660	5.32
14	suburban	103	3.69	alien	8	3.20	overall	592	3.63
15	surgical	98	4.26	coastal	8	3.28	elderly	456	4.40
16	downtown	93	3.06	said	8	3.82	increasing	433	3.68
17	nursing	90	4.10	scattered	8	4.93	ethnic	420	3.55
18	full-time	81	3.36	impoverished	8	5.03	patient	403	4.11
19	uninsured	80	5.02	surviving	7	4.32	homeless	378	4.09
20	eligible	75	3.05	occupied	7	5.19	genetic	368	3.59
21	displaced	71	5.50	isolated	6	3.12	rapid	325	3.86
22	foreign-born	69	6.78	fictional	6	3.54	representative	258	3.99
23	inner-city	64	5.41	primitive	6	3.72	working	249	3.06
24	out-of-state	62	5.86	present-day	6	5.16	declining	220	5.18
25	surrounding	61	3.22	elusive	6	4.15	estimated	213	3.55
26	undocumented	58	4.49	neighboring	5	3.03	distinct	201	3.15
27	unempl	56	3.44	non-	4	6.83	underse	187	7.04

	oyed			jewish			rved		
28	affected	52	4.39	humanoid	4	6.72	demographic	178	3.52
29	first-year	51	5.16	sentient	4	5.34	viable	171	3.69
30	part-time	46	3.23	unwelcome	4	4.85	isolated	169	3.83

Table 2 shows the most common 30 adjective collocates used with the lemmas *resident, inhabitant,* and *population* on the basis of COCA data. Each item features the frequency of the collocate and its MI score (Mutual Information Score), which quantifies the intensity of the collocational association. If we examine the data, we can easily discern patterns when it comes to the use of each of these three highly similar nouns in a wide range of different situations. The noun *resident* is most commonly found with adjectives of legal status, permanence, or location.

The high-frequency collocations are local (2,034), longtime (784), permanent (779), and legal (612), suggesting that the word occurs in talk about persons who are *resident* in a place for extended or indefinite periods, often with a legal or civic implication. Other such collocations as urban, low-income, lawful, and year-round support the hypothesis that "resident" is used frequently in government, social, or housing contexts. The best-scoring collocate is foreign-born (6.78), with a very strong but rare co-occurrence, likely to be encountered in discussion of immigration or census data. It is paired with *inhabitant,* used in more historical, anthropological, or denotive situations of ancient or native *populations*. While frequencies are generally lower overall (e.g., local only appears 151 times with *inhabitant*), the MI scores are high, a sign of strong specific connections.

The collocates native, indigenous, aboriginal, ancient, and prehistoric suggest the typical use of the word to describe early people or groups of people closely linked with territory or tradition. High MI values for words such as non-Jewish (6.83) and humanoid (6.72) suggest use in scientific or academic writing, particularly in anthropology, archaeology, or history. *Population,* on the other hand, is most closely

linked with demographic, statistical, and social environments. Its most frequent collocates are large (4,427), general (4,024), total (2,033), and whole (1,758), which indicate that it is employed in quantitative or overall description of groups of people. Collocates like growing, vulnerable, aging, and urban reveal that "population" is often discussed in processes of change, vulnerability, or needs.

Whereas frequency is more inclusive, certain collocates like underserved (MI score 7.04) and indigenous (5.32) imply strong, precise association with minority or marginalized people. In brief, while *inhabitant* and *population* are near synonyms to *resident*, they collocate with strongly different sets of adjectives that reflect their typical use environments. *Resident* is used in civic and legal environments, *inhabitant* in histories or anthropologies, and *population* in demographic or policy environments. The evidence shows the manner in which word choice subtly influences meaning and tone even between similar-definition words.

4.3 Semantic Preference

4.3.1 Semantic Preference of Resident

Table 3

Semantic Preference of Resident

Semantic Themes	Adjective Collocates
Geographic	local, nearby, coastal, rural, urban, suburban, downtown, inner-city, surrounding
Duration of Residence	longtime, lifelong, permanent, year-round, first-year
Legal Status and Immigration Status	legal, lawful, undocumented, foreign-born, out-of-state, displaced
Employment and Education Status	full-time, part-time, unemployed, eligible, first-year

Socioeconomic Status	low-income, uninsured, affected
Demographic	elderly, nursing
Medical and Professional Context	surgical, first-year , full-time

The semantic preference of the noun *resident*, as revealed through its adjective collocates, reflects subtle usage patterns that highlight its association with a wide range of socio-geographical and institutional contexts. Among the most salient semantic domains is the geographical theme. Collocates such as local, nearby, coastal, rural, urban, suburban, downtown, inner-city, and surrounding demonstrate how *resident* is used again and again to describe physical or locational properties. They function to place the *resident* in space in a circumscribed area and are found typically in texts related to regional planning, neighborhood statistics, or public policy. Phrases such as urban *residents* or coastal *residents* are generally employed in writing pertaining to environmental planning, socio-economic development, or disaster responses. This signifies the close association between *resident* and spatial definition in administrative or civic discourse. Another prominent theme involves residence duration.

Adjectives such as longtime, lifelong, permanent, year-round, and first-year suggest that *resident* is normally encountered in situations where the temporal nature of an individual's residence is important. These adjectives not only indicate how long a person has been living somewhere but further suggest some sort of attachment or official status. In particular, permanent *resident* is an immigration and civil law term, while first-year *resident* tends to appear typically in educational or medical settings, referring to new entrants or trainees. This temporal aspect is typically key to the storylines around community integration, residential rights, and institutional roles. Legal and immigration status is likewise a significant theme that clearly characterizes the semantic profile of *resident*. Collocates such as legal, lawful, undocumented, foreign-born, out-of-state, and displaced create the extent to which the term is rooted in discourses of legality, citizenship, and mobility.

These adjectives are employed both to mark where someone resides and the position which such a person is in in relation to formal institutions and state authorities. As an example, foreign-born and undocumented residents are often used in policy and legal documents on immigration reform, access to public goods and services, and social integration. Such collocational patterns contribute to the reading that *resident* tends to call forth legal or administrative connotations over its denotative connotation. Employment and education contexts also impact the use of *resident*, as seen through collocates full-time, part-time, unemployed, eligible, and first-year. These adjectives are least frequently used in casual, academic, or official composition and are especially used in contexts in which individuals are categorized by functional position. In higher education and health care, *resident* is routinely referred to describe individuals who are undergoing postgraduate training, e.g., first-year *residents* of medical residency programs.

Phrases such as *unemployed residents* or *eligible residents* are used more frequently than average in discussions related to labor force participation, employment training, and access to welfare services. This reflects the flexibility of the term in the explanation of socioeconomic participation in an organized system. At the socioeconomic level, the collocates low-income, uninsured, and affected reflect *resident* being most frequently used to describe vulnerable or marginalized groups. Such collocations occur primarily in government or public health discourse where populations are categorized by need, access, or risk. For instance, low-income *residents* can be targeted for the housing subsidy, and uninsured *residents* are those evoked in healthcare reform discussions. These trends point to how *resident* can index social inequality or need-based categorizations in administrative language.

The word *resident* also occurs frequently in demographic categories, most frequently with adjectives such as elderly and nursing. These combinations are common within healthcare, senior services, or residential care facility contexts. They suggest that *resident* is used to describe populations based on age or care classification, such as elderly residents of assisted living facilities or nursing home residents. These uses justify the pervasiveness of the word across demographic and social service writings. Finally, in professional and medical contexts, collocates surgical and first-year echo institutionally the use of resident in postgraduate medical training. They typically

occupy formal occupational titles, since the term surgical resident, which refers to trainees in organized, hierarchical professional systems. This is a case of specialized use of the term within medicine, particularly in North American settings, where the term resident refers to a specific position with clearly defined duties and position. In short, the semantic variation documented through these adjective collocates confirms that *resident* is an extremely polysemous word with varied functions in formal registers. It is used in legal, medical, geographical, and socioeconomic usage, frequently in institutionally determined senses. These tendencies validate the contextual flexibility of *resident*, as well as its central position in the expression of individual identity in terms of systems of place, law, provision of services, and structures of profession.

4.3.2 Semantic Preference of Inhabitant

Table 4

Semantic Preference of Inhabitant

Semantic Themes	Adjective Collocates
Geographic	local, rural, coastal, surrounding, neighboring, isolated, occupied
Origin, Ethnicity, or Identity	native, indigenous, aboriginal, original, non-jewish
Time-Based Descriptor	ancient, prehistoric, primitive, present-day, remaining, said
Current State or Condition	peaceful, fellow, scattered, impoverished, elusive, unwelcome, surviving, sole
Non-Human or Fiction	alien, fictional, humanoid, sentient

The semantic profile of the noun *inhabitant* reveals a narrower and more contextually specialized usage compared to *resident*, with strong associations to geographic, ethnographic, historical, and fictional themes. One of the most prominent

semantic domains in which *inhabitant* is used is geographical or locational. Adjectives such as local, rural, coastal, surrounding, neighboring, isolated, and occupied suggest that *inhabitant* is often employed to describe people in relation to specific physical spaces. These collocates indicate that the term tends to occur in descriptive or observational contexts, often focusing on regions that are distinct or remote. For instance, isolated *inhabitants* and coastal *inhabitants* may appear in travel writing, historical descriptions, or environmental studies where emphasis is placed on geography and spatial orientation.

A second semantic theme that consistently appears in the collocational data is origin, ethnicity, or identity. Adjectives such as native, indigenous, aboriginal, original, and non-Jewish point to inhabitant being used in texts that discuss heritage, ancestral connection, or cultural identity. These adjectives indicate that inhabitant is more likely than *resident* to describe people in collective terms that foreground lineage or group affiliation, especially in anthropological or sociopolitical discourse. For example, indigenous *inhabitants* might be referenced in discussions of colonialism, land rights, or traditional ecological knowledge. The lexical pattern here suggests a preference for historical and ethnic framing, emphasizing continuity, dispossession, or cultural resilience.

Closely related is the time-based semantic category, which comprises adjectives such as ancient, prehistoric, primitive, present-day, remaining, surviving, and sole. These adjectives place *inhabitant* in a diachronic context, distinguishing it from more contemporary or bureaucratic terms like resident. The use of prehistoric *inhabitants*, for instance, is common in archaeological or anthropological writing, whereas surviving *inhabitants* may appear in historical narratives describing remnants of past civilizations or groups impacted by war, disease, or displacement. This temporal anchoring contributes to the term's formal and literary tone, emphasizing its descriptive and referential nature rather than practical or legal implications.

Besides time and ethnicity, *inhabitant* is also characterized by a variety of adjectives conveying social, emotional, or status meaning. Among them are peaceful, fellow, scattered, impoverished, elusive, and unwelcome. The collocates show that *inhabitant* is able to be employed in narratives where social conditions, states of mind, or marginality are at issue. For example, poor *residents* might be mentioned in

sociological discourse, elusive or unwanted *residents* in fictional or metaphorical domains. This is an indication of the tendency for the term to be employed evaluatively, habitually charging it with connotations of strife, displacement, and otherness.

Another interesting aspect of *inhabitant*'s semantic behavior is that it is associated with fictional, non-human, or imaginary referents. Adjectives such as alien, fictional, humanoid, and sentient express the usual sense of the word in science fiction, science fantasy, or mythological prose. These usages suggest that *inhabitant* not only describes actual human beings, but can simultaneously hold space for imaginary or symbolic beings. This term reinforces its literary or creative aspect and differentiates it from the more specific, administrative term *resident*.

Overall, the adjectival collocates of *inhabitant* position the word within a lexical and stylistic territory that tends towards formal, literary, historical, and speculative registers. Its usual collocations are with ethnicity, antiquity, isolation, and fictionalization, which suggest that *inhabitant* is used relatively infrequently in functional or administrative discourse and more usually in writing intended to describe, narrate, or contextualize populations in terms of identity, geography, or temporality. As such, *inhabitant* has a unique role in English lexical semantics as a word with high descriptive and symbolic potential but low conversational and institutional use.

4.3.3 Semantic Preference of Population

Table 5

Semantic Preference of Population

Semantic Themes	Adjective Collocates
Size and Scope	large, total, entire, overall, estimated, viable, general
Change and Growth Trends	growing, increasing, declining, rapid
Geographic	urban, rural, isolated

Ethnicity or Origin	native, indigenous, ethnic, genetic, distinct, diverse
Demographic and Social Characteristics	civilian, working, elderly, aging, homeless, patient, vulnerable, underserved, representative, demographic

Table 5, showing the semantic preferences of the noun *population*, as revealed by its collocating adjectives, illustrates that the term is employed virtually only in formal, academic, and data-driven discourse. The most salient semantic field to be linked with *population* relates to size and scope. Adjectives such as large, total, entire, overall, estimated, and viable indicate the noun's frequent use in quantitative and statistical contexts. These modifiers emphasize magnitude, totality, or proximity, and are common in texts dealing with national censuses, ecological sustainability, or resource allocation. Total *population* or sustainable *population* of a species are commonplace in demographic studies or environmental impact assessments. This lexical preference reinforces *population* as a collective noun for enumerating or describing groups from an aggregated level rather than on a per capita basis. Closely connected with this is the issue of change or growth trends.

Such words as growing, increasing, declining, and rapid tell us that *population* is frequently used in texts which involve temporal change, development, or crisis. These collocates reflect the use of the noun in dynamic systems, namely scientific ones such as sociology, economics, epidemiology, and town planning. For instance, the phrase declining *population* and rapid *population* growth are recurring in academic papers that discuss birth rates, migration, or stresses on social infrastructure. This combination sets out the word's scope for analysis and priority in use to define trends in society. The locational or geographic theme also reveals *population* utilized in categorizing groups based on location. Urban, rural, and remote are adjectives that indicate how the word is generally used in regional comparisons or where service accessibility and development differentials are concerned. Urban *population* and rural *population*, for instance, are

continuously compared in studies evaluating policy impact, educational achievement, or health service availability.

The word's permissiveness in delineating place-based *populations* provides it with a useful tool for use in sociopolitical analysis and institutional rhetoric. A second wide semantic field is ethnicity and origin, with native, indigenous, ethnic, genetic, distinct, and diverse as collocates. These adjectives suggest that *population* is often employed to characterize groups in terms of cultural origins, ancestry, or biological characteristics. In areas such as public health, genetics, or multicultural policy, *population* is a comparative analysis term for between-identity-groups. Phrases such as ethnic *population* or genetic *population* are used in those fields and are used to highlight the abstractness and analytical merits of the term.

Finally, the subject of demographic and social characteristics can be found embedded in adjectives such as civilian, working, elderly, aging, homeless, patient, vulnerable, underserved, representative, and demographic. This set of collocates implies the noun's role as a primary categorizing criterion of individuals in terms of social role, health, or service need. *Population* is also commonly employed to refer to target *populations* in social policy and public health, especially in terms like vulnerable *population*, underserved *population*, or aging *population*. These usages reflect how the term is utilized as a means of defining, as well as addressing, social problems collectively.

4.4 Comparison of Themes between Resident, Inhabitant, and Population

Table 6

Comparison of Themes between Resident, Inhabitant, and Population

	Resident	*Inhabitant*	*Population*
Geographic	✓	✓	✓
Demographic	✓	✗	✓
Ethnicity	✗	✓	✓

The table gives a comparative picture of semantic themes attested symmetrically by the near-synonymous nouns' adjective collocates, *resident*, *inhabitant*, and *population*. Specifically, it illustrates three main themes—geographic, demographic, and ethnic—and whether or not each of the target words co-occurs with adjective collocates expressive of those semantic domains.

Each of the three words—*resident*, *inhabitant*, and *population*—is shown to share the geographical thread, indicated by check marks in row one. This means that each of these words is likely to be used in contexts where location, region, or spatial belonging is prominent. For instance, all three words share collocates such as urban, rural, coastal, and local, which suggests a shared semantic bias to associate human occupation with physical or locational characteristics. This shared fondness for locational words points up their straightforward denotative identity.

As contrasted with this, the demographic theme—traditionally applied to group categories such as old, working, or vulnerable—is more closely shared. It is clearly associated with *resident* and *population* but not with *inhabitant*. This is to say that *resident* and *population* are usually used in discussion evoking demographic categorization, mostly in policy, health, or social welfare contexts. For example, phrases such as aging *population* or low-income *residents* are common in institutional writing concerned with resource distribution or community health. The absence of this theme for *inhabitant* means that the term is used less often for *population* or policy purposes, and instead is employed in more general or descriptive or historical contexts.

The final theme, ethnicity, appears in both *inhabitant* and *population*, but not in *resident*. This shows that *population* and *inhabitant* tend to co-occur with such words as aboriginal, indigenous, ethnic, and native, most of which tend to be used in group identity, ancestry, or cultural origin. The occurrence of this topic in both words suggests their common use in academic, anthropological, or sociopolitical contexts where identity and lineage are key. The omission of this theme by *resident* ensures that the term is generally more neutral and institutional, usually without cultural or ethnic overtones, and principally used in legal, medical, or administrative texts.

Finally, while the three words share an essential referential function, their semantic tendencies are significantly different. *Resident* and *population* are more suitable in modern, data-driven, and policy-driven contexts, particularly in

demographic classification, whereas *inhabitant* has a more descriptive, antique, and ethnographic bias. The intersection of geographical topics confirms their underlying semantic core, but the difference in demographic and ethnic meanings testifies to their diverse discursive functions.

CHAPTER 5
DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary of the Findings

5.1.1 Frequency and Distribution Across Genres

The frequency analysis retrieved from the Corpus of Contemporary American English (COCA) revealed a substantial disparity in the overall use of the three target words. Among them, *population* occurred most frequently (111,449 tokens), followed by *resident* (20,716 tokens) and *inhabitant* (522 tokens). These findings support Edmonds and Hirst's (2002) claim that near-synonyms are often differentiated by their pragmatic distribution rather than core semantic meaning. Although all three terms refer to people in a given area, the scale, tone, and discourse function differ considerably.

In terms of genre distribution, *population* appeared predominantly in academic texts, followed by magazines and news sources. This distribution affirms its high formality and technical application, especially in contexts involving statistical analysis, demographic studies, or policy reports. As predicted in Chapter 3, *population* functions largely as a collective noun that refers to groups of people viewed quantitatively, which is typical of scientific or institutional discourse.

Resident, in contrast, was more frequent in news articles, web content, and spoken texts, suggesting a broader range of usage across both formal and informal settings. Its presence in spoken discourse (17.79 per million words) indicates a relatively high degree of accessibility and familiarity. This aligns with findings from Sinclair (1991) and Phoocharoensil (2010), who observed that collocational and genre distributions reflect lexical naturalness and domain specificity.

Inhabitant occurred least frequently and appeared primarily in academic and fiction genres. Its extremely low frequency in speech (eight tokens) points to its highly formal or literary character. These observations confirm Cruse's (1986) argument that near-synonyms often differ in terms of register, even when they share a similar denotative core.

5.1.2 Collocational Patterns

Analysis of high Mutual Information (MI) score collocates offers deeper insights into how the three terms pattern with surrounding words. The word *resident* commonly collocates with descriptors such as *local*, *foreign*, *legal*, and *permanent*, all of which reflect legal, geographical, or administrative connotations. These collocates suggest that the term *resident* is frequently used in civic, medical, and policy-related contexts—e.g., *legal resident*, *hospital resident*, and *foreign resident*. Such usage reflects the notion of "lexical priming" (Hoey, 2005), in which words are predisposed to occur with certain lexical items based on habitual usage patterns.

In contrast, *inhabitant* frequently co-occurs with collocates such as *native*, *indigenous*, *aboriginal*, and *ancient*. These adjectives reflect ethnic, historical, or geographical associations, marking the term as more suitable for descriptive, anthropological, or literary contexts. Notably, some collocates such as *humanoid* or *non-human* also suggest usage in speculative or fictional registers, further limiting the word's range in everyday usage.

Population, on the other hand, collocates strongly with adjectives indicating scope and condition—e.g., *aging*, *vulnerable*, *total*, and *urban*. These pairings are common in demographic discourse and institutional planning. As Murphy (2003) noted, such collocational behavior is not random but shaped by the lexical and semantic environments in which the word is regularly used. The collective and impersonal nature of *population* is thus reinforced by its collocates, which reflect aggregation and social categorization rather than individuality.

5.1.3 Semantic Preference

Semantic preference, defined as the regular co-occurrence of a lexical item with words sharing a semantic field (Stubbs, 2001), further illustrates the divergence among the three synonyms. *Resident* aligns with themes such as legal status, institutional belonging, and community membership, making it the most suitable term for use in modern administrative and medical contexts.

Inhabitant, however, exhibits a strong preference for themes related to ethnicity, primordial belonging, and territorial origin. Its typical collocates mark it as archaic or descriptive, often detached from present-day realities. This supports Cruse's (1986)

notion of "expressive variation" among near-synonyms, where words differ in stylistic effect despite overlapping meanings.

Conversely, *population* consistently aligns with themes such as quantification, vulnerability, and demographic change. Its usage in contexts such as *urban population, aging population*, and *underserved population* shows that it is frequently employed to refer to social groups in planning or evaluative contexts. These semantic patterns illustrate that although the three terms are synonymous in dictionary definitions, their meanings diverge significantly in actual use due to semantic preference and register.

5.2 Conclusion

This study gives certain and complete answers to both research questions, bringing empirical support through corpus-based analysis to distinguish between the near-synonyms *resident, inhabitant,* and *population*. The first research question sought to examine variation of the three synonyms in terms of frequency and distribution across genres. The research found that *population* is the most common word utilized, particularly in official and technical writing such as in academic journals, magazines, and news reports. Its regularity in these forms of writing testifies to its frequent application in institutional, statistical, and demographic uses. *Resident*, while significantly less frequent than *population*, showed consistent use across both formal and semi-formal genres, including news, spoken texts, and web content, indicating its flexibility and relevance in civic and legal domains. Conversely, *inhabitant* was the least frequent and most restricted in usage, occurring predominantly in academic and fictional genres. Its near absence in spoken texts and institutional discourse further emphasizes its literary or descriptive function. These findings suggest that although the three words share a similar denotative core, their degree of formality, register, and usage domains differ substantially, thereby answering Research Question 1.

The second research question focused on the typical collocates—both adjectives and nouns—of each synonym, as well as their semantic preferences. The collocational analysis provided strong evidence of how each word is used in distinctive contexts. *Resident* co-occurred with collocates such as local, permanent, foreign, and legal, highlighting its association with legal status, healthcare, and residential governance. This supports the notion that *resident* often refers to individuals with legal or civic

recognition within a community. *Inhabitant*, on the other hand, collocated with adjectives like native, aboriginal, ancient, and indigenous, suggesting its typical use in ethnographic, historical, or geographical narratives.

These collocates also demonstrate that *inhabitant* is less commonly used in contemporary or conversational contexts. *Population* most frequently co-occurred with the adjectives aging, vulnerable, urban, and underserved, which confirmed its semantic correspondence with public policy language and large-scale demographic classification. Each word's semantic preferences also supported these collocation patterns. *Resident* was linked to themes of institutional affiliation and civic identity, inhabitant was associated with themes of historic presence and ethnicity, and population was linked to quantification and collective identity. These findings conclusively address Research Question 2 in that they demonstrate that all of the synonyms exhibit unique collocational tendencies and semantic preferences that restrict their substitutability.

As a group, the findings strongly establish that *resident, inhabitant,* and *population* are near-synonyms and not strict synonyms. Their choice is highly susceptible to genre, collocational context, and discourse use, thereby upholding the theory accounted for in Chapter 2. The findings also uphold the worth of contextual fit in instructing vocabulary and separating out synonyms, thereby vindicating the original aims and motivation of this study.

5.3 Pedagogical Implications

The findings of this study have significant implications for English Language Teaching (ELT), particularly in the areas of vocabulary instruction and academic writing. Non-native learners often struggle to select the appropriate synonym due to limited understanding of register, collocation, and semantic nuance (Laufer, 1990; Webb, 2007). Errors such as *She is a population of this town* illustrate the issue of incorrectly substituting near-synonyms based solely on dictionary definitions, without accounting for contextual and grammatical appropriateness.

In response to this issue, ELT instructors must incorporate corpus evidence in educational materials. Concordance lines from COCA can be used to illustrate the students that each word functions differently depending on genres and contexts. For

example, the application of *resident* in medical or law papers, *inhabitant* in chronicles, and population in statistics documents can be illustrated to students. Such use-based instruction aligns with data-driven learning practice (Gilquin, 2020), which has been identified to make students more sensitive to collocation and register.

In addition, structured vocabulary activities such as synonym substitution drills, genre-matching drills, or collocation sorting exercises can help learners acquire these contrasts. Teachers can also draw attention to collocational profiles and semantic collocates by getting students to observe the contrast between structures such as aging *population* and foreign *resident*, thereby sensitizing learners to context-dependent lexical choice.

In general, sensitivity to contextual constraints and traditional uses of near-synonyms contributes to lexical accuracy, stylistic appropriateness, and communicative competence—goals especially important to learners who want to use English academically or professionally.

5.4 Limitations

Although this study yielded useful findings despite the constraints, some must be mentioned. To begin with, the study utilized the Corpus of Contemporary American English (COCA) only, which, as broad and balanced as it was in genres, includes primarily American English. As such, the findings may not necessarily reflect synonym use in other forms of English, e.g., British, Australian, or Indian English. Future research may consider using the British National Corpus (BNC) or the International Corpus of English (ICE) data for comparison. Second, only three target synonyms were analyzed. While this allowed for close and detailed examination, it limited the scope of generalizability. Other near-synonyms for groups of people or place identities—e.g., citizen, dweller, or denizen—could yield further insights if included in future studies.

Third, this study primarily tested adjective and noun collocates using MI scores to investigate semantic preference. Although this process is standard and correct, it does not uncover more advanced grammatical or prosodic structures such as syntactic frames or semantic prosody. Further investigation of verbs, syntactic patterns, or discourse analysis could give us a fuller view of synonym behavior.

Finally, while this research offers pedagogical implications, it does not empirically test the influence of corpus-based instruction on learners' use or understanding of near-synonyms. A follow-up study with a classroom experiment, learner feedback, or pre/post-testing would increase practical applications of these findings in English language teaching.

Recognizing these constraints serves to situate the present work as a precursor to further studies, and highlights the necessity for more extensive and subtle examinations into synonymy, lexicon acquisition, and corpus-based pedagogy.

REFERENCES

Aroonmanakun, W. (2015). *A corpus-driven study of the adjectives quick and fast*. *Journal of Language and Linguistics,* 33(2), 45–58.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Cai, H. (n.d.). *A corpus-based study of synonymous adjectives denoting "very good"*. [Unpublished manuscript].

Charles, M. (2004). The construction of stance: A corpus-based investigation of two contrasting disciplines. *Journal of English for Academic Purposes*, *3*(4), 313–326. https://doi.org/10.1016/j.jeap.2004.07.002

Chung, S. (n.d.). *A corpus-based study of near-synonyms: Create and produce*. [Unpublished manuscript].

Davies, M. (2008–). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

Davies, M. (2011). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, *25*(4), 447–464. https://doi.org/10.1093/llc/fqq018

Divjak, D., & Gries, S. T. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, *2*(1), 23–60. https://doi.org/10.1515/CLLT.2006.002

Flowerdew, L. (2012). *Corpus-based language studies: An advanced resource book*. Routledge.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Jackson, H. (1988). *Words and their meaning*. Longman.

Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.

Krawczak, M. (2014). Synonyms in context: A corpus-based study of adjectives denoting shame. *Research in Language, 12*(4), 355–367. https://doi.org/10.2478/rela-2014-0019

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh

University Press.

Mindt, D. (1998). An empirical grammar of the English verb: Modal verbs. In S. Granger (Ed.), *Learner English on computer* (pp. 121–137). Longman.

Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*. Cambridge University Press.

Palmer, F. R. (1997). *Semantics* (2nd ed.). Cambridge University Press.

Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. John Benjamins Publishing.

Petcharat, S. (2016). *A corpus-based study of the near-synonymous adjectives appropriate, proper, and suitable* [Master's thesis, Thammasat University].

Phoocharoensil, S. (2010). A corpus-based study of English synonyms: Ask, beg, plead, request and appeal. *Theory and Practice in Language Studies, 1*(2), 151–160. https://doi.org/10.4304/tpls.1.2.151-160

Sakulratanacharoen, R. (2013). *A corpus-based analysis of English reporting verbs in news discourse* [Master's thesis, Chulalongkorn University].

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.

Thamratana, T. (2013). *A corpus-based analysis of synonyms related to reduction in English* [Master's thesis, Chulalongkorn University].

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins Publishing.