

Developing Resources for the Study of Phraseological Sophistication

University Name: **Catholic University of Louvain**

Author: **Jiacheng Shen**

Promoter: **Dr. Magali Paquot**

Academic Year: **2022-2023**

Major and Specialization: **Linguistics. Linguistics and its Application for
a Multilingual Society**

Table of Contents

1. Introduction.....	3
2. Phraseology and Second Language Research.....	6
2.1 Defining Criteria	7
2.1.1 Polylexicality	7
2.1.2 Compositionality.....	8
2.1.3 Restricted Collocability	13
2.2 Categorization	14
2.2.1 The phraseological approach	15
2.2.2 The phraseological approach	18
2.3 Zooming in on Collocations.....	20
2.3.1 The phraseological approach	20
2.3.2 The distributional approach	25
3 Phraseology and Complexity in L2 Language.....	29
3.1 Studies of Multiword Units in Learner Language.....	29
3.2 Studies of Phraseological Complexity	33
4. Multiword Unit Lists Construction.....	40
4.1 Multiword unit lists constructed in previous research	40
4.2 Constructs and Operationalizations.....	47
5. Data.....	54
5.1 Data Description.....	54
5.2 Data Pre-processing.....	59
5.3 Constructs for this Study	62
5.3.1 Frequency.....	63
5.3.2 Dispersion	63
6. Developing the New Academic Collocation List.....	65
6.1 Filtering method	65
6.2 Result.....	68
6.2.1 Composition.....	68
6.2.2 Coverage	69
7. Discussion.....	72

8. Conclusion	97
9. References	99
10. Appendix	106

1. Introduction

The acquisition, fluency, and competency of a language have been demonstrated to be significantly impacted by the use of various types of word combinations, including idioms, restricted collocations, proverbs, fixed expressions, and similes (e.g., Coxhead, 2008). In the domain of learner corpus research, the importance of phraseological competence for achieving fluent and idiomatic language use has been widely acknowledged (Paquot, 2018) and numerous researchers conducted various research to investigate how learners use such word combinations. For instance, the study conducted by Durrant and Schmitt (2009) revealed that non-native writers of English heavily relied on high-frequency collocations, but they exhibited a tendency to underuse less frequent, strongly associated collocations. Similarly, another study conducted by Siyanova-Chanturia and Schmitt (2008) focused on adjective + noun collocations in the English writings of Russian learners and suggested that despite their ability to produce a significant number of frequent and strongly associated English word combinations, even advanced learners' underlying intuitions and fluency with collocations did not match those of native speakers. These findings were further supported by Bestgen and Granger's (2018) longitudinal investigation of the phraseological development of English as Foreign Language (EFL) learners, which demonstrated an increase in the proportion of collocations with high pointwise mutual information (PMI) scores produced by learners over time. Collectively, these studies highlight that learners of different proficiency levels may exhibit varying degrees of mastery in phraseology, and the use of phraseology can serve as a useful indicator for examining L2 performance.

Recent research has focused on examining the complexity of L2 performance in terms of phraseological complexity, which Paquot (2019:126) referred to as “the range and degree of sophistication of phraseological units used in language production”. In Paquot (2019),

phraseological units were operationalized as binary dependency relationships between a head and its dependent. The operationalization of phraseological complexity included two dimensions: phraseological diversity and phraseological sophistication, which respectively capture the breadth and depth of phraseology. Three types of dependency relations (i.e., adjectival modifiers, adverbial modifiers, and direct objects) were used to operationalize phraseological complexity. Phraseological diversity was operationalized as the root type-token ratio of the three dependencies. Phraseological sophistication was assessed using two methods. The first method employed the mean pointwise mutual information (PMI) score, while the second method involved the calculation of type- and token-based ratios of “sophisticated collocations”. These ratios were calculated by comparing the quantity of sophisticated dependencies to the total number of dependencies. The study used a corpus of linguistics essays written by L1 French EFL learners. The result of employing the second methodology to measure phraseological sophistication unveiled that although the ratios of phraseological sophistication increased with learners' overall proficiency levels according to the Common European Framework of Reference (CEFR; Council of Europe, 2001), no statistically significant differences were observed.

Research on phraseological sophistication within second language (L2) writing, operationalized through the utilization of both type- and token-based ratios derived from the division of the count of sophisticated dependencies by the total count of dependencies, as exemplified in the investigations conducted by Paquot (2018, 2019), has employed Ackermann and Chen's (2013) Academic Collocation List (ACL) to operationalize the dimension of phraseological sophistication. Findings have revealed that the more proficient L2 academic writers are, a larger number of academic collocations included in the ACL are employed in their writings, but these differences have not been statistically significant. This could be attributed to the very nature of the ACL. The ACL was designed to assist EFL learners in improving their

collocational competence and to aid English for Academic Purposes (EAP) teachers in lesson planning. However, it was not designed for research purposes to examine L2 EFL learners' phraseological competence. As a result, the list is relatively small and has limited coverage in its source corpus, with the academic collocations only accounting for 1.4% of the source texts used in Ackermann and Chen's study. Therefore, its use as a measure of phraseological sophistication is limited.

Hence, the primary objective of this research is to create a new academic collocation list. This list is expressly designed for research purposes, serving as a quantitative gauge for measuring phraseological sophistication of L2 EFL writings. This list departs from its conventional role as a pedagogical tool for EAP teachers and students. The intended development of this new list stems from the aspiration to address two issues encountered within the ACL: its size and coverage. The new list, on one hand, will encompass an amplified assortment of collocations that are inherently characteristic of academic genre. Concurrently, it will exhibit an expanded coverage across a broader array of academic texts from which the list shall be extracted. This way, the list will be a better reflection of the various kinds of academic collocations that are representative of academic genres across academic disciplines.

In its ultimate form, this novel list will serve as a potent resource for delving into the intricacies of phraseological sophistication within L2 EFL writings. It is poised to facilitate nuanced explorations into the linguistic nuances prevalent across diverse academic disciplines. Furthermore, it will help researchers to better understand the language used in academic writing across different disciplines, and to evaluate the language proficiency of learners in these contexts.

2. Phraseology and Second Language Research

According to Knappe (2004), phraseology emerged as an academic method for studying language in the 20th century. Initially, the study of phraseology began in the former Soviet Union and other Eastern European countries (Cowie, 1998). In recent years, there has been a significant surge in interest regarding the field of phraseology, which includes both the study of the phenomenon itself and its intersection with other disciplines such as learner corpus research. Recently, phraseology has come to an intersection of various linguistic disciplines, including lexicology (De Cook, 2000), psycholinguistics and language acquisition (Wray, 2002), second language acquisition (Nation, 2001), and English for academic purposes (Cowie, 1997).

Cowie (1994) provided a general definition of phraseology as the study of the structure, meaning, and use of word combinations, leaving much room for ambiguity as mere explanation or specification has been made for the concept “word combinations”. Researchers have proposed various terms to refer to the object of study, including word-combinations (Cowie, 1994), multi-word units (De Cock, 2003), fixed expressions (Alexander, 1984), and formulaic sequences (Wray, 2002). Despite the many proposed definitions, there is still no clear agreement on the definition of phraseology, with some definitions being opaque or conflating different terms (Gries, 2008).

Gries (2008:2) proposed a contemporary definition for phraseologism or phraseological units, wherein they are described as “the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance”.

The purpose of this study is to provide collocational resources for evaluating the phraseological sophistication of EFL learners. Therefore, it is crucial to thoroughly review the existing literature on phraseology, a broader term that encompasses collocations. The subsequent sections present a comprehensive examination of diverse aspects of phraseology. Section 2.1 adopts a specific perspective by focusing on several defining criteria for phraseology, with the aim of delving into its fundamental nature. In Section 2.2, various categorizations of phraseology proposed by prior researchers are explored. Shifting the perspective to collocations, which constitute the primary focus of this research, Section 2.3 seeks to uncover the nature of collocations and presents the definition of ‘collocation’ in this study.

2.1 Defining Criteria

Due to the diverse range of approaches taken by researchers in exploring and defining phraseology, arriving at a universal definition of the field has proven challenging. Nevertheless, several shared criteria have been identified and examined by scholars. This section will scrutinize several such criteria, including polylexicality, fixedness, and collocability.

2.1.1 Polylexicality

Polylexicality in phraseology generally refers to the fact that multiword units are made up of more than one lexical item (i.e., a word or a group of words with a lexical meaning).

Polylexicality is a distinguishing feature of all types of multiword expressions, including idioms, collocations, and phrasal verbs.

For example, the idiom *spill the beans* is a polylexical unit, as it is made up of the verb *spill* and the noun *beans*. Similarly, the collocation *read a book* is a polylexical expression, as it is made up of the verb *read* and the noun *book*.

Polylexicality is a key characteristic of phraseology. The fact that multiword units are made up of more than one lexical item constitutes a basic characteristic of phraseology.

2.1.2 Compositionality

Barkema (1996:138) provided a definition of compositionality as “the degree to which the interpretation of a multiword unit can be determined by the combination of the basic or derived meanings of the lexical items contained within the unit and the syntactic relationships within the constituent that encompasses these lexical items”. Prior to introducing this concept, Barkema first differentiated between three types of senses for a lexical item, e.g., basic sense, extended sense, and derived sense.

The concept of lexical sense can be categorized into three types, as identified by Barkema (1996:138). The first type is the basic sense, which refers to the primary and most salient meaning of a lexical item that comes to mind when it is encountered as a standalone unit, rather than within a larger sentence. This sense forms the foundation for all other senses that may be associated with the lexical item. The second type is the extended sense, which arises as a result of violating a restriction rule in a specific context, leading to an expansion of the fundamental meaning. The third type is the derived sense, which denotes a secondary meaning of a lexical item that has become established in the speaker's mind. Typically, this sense is the second or third interpretation that comes to mind when the speaker encounters the lexical item as a single word. The subsequent paragraph presents a collection of illustrative instances concerning the three distinct types of senses.

The basic sense of the lexical term *house* typically refers to a structure intended for human habitation. However, the extended sense of *house* is exemplified in the sentence "But most Czech believed until now that Slovakia would still be in favor of a federal house" (The Economist, 1/8/1992). Here, the term *house* transcends its literal meaning and refers to a *country*.

Furthermore, the derivational sense of *house* is illustrated in the sentence "Sir Marcus divided the house and did not divide without organization. Hughes's motion was roundly defeated" (New Statesman & Society, 2/8/1991). According to the Oxford English Dictionary, in this context, *house* signifies *a group of people who meet to discuss and make the laws of a country*.

The concept of compositionality was approached via such three senses of a lexical item. In other words, the meaning of a multiword unit is compositional if it can be predicted based on the meanings of its constituent words.

For example, the meaning of the multiword unit *red apple* can be derived from the meanings of the individual words *red* and *apple*. The meaning of the multiword unit is compositional because it reflects the combination of the meanings of the two words. However, not all multiword units are compositional. Some multiword units have a meaning that cannot be predicted from the meanings of their individual words. For example, the multiword unit *kick the bucket* means to die, but this meaning cannot be derived from the meanings of the words *kick* and *bucket* alone.

Barkema (1996) developed a classification scheme that distinguishes between different levels of compositionality in language, which can be categorized into four levels: fully compositional, pseudo-compositional, fully non-compositional, and partly non-compositional. These levels are elaborated upon as follows:

- Fully compositional constructions are those in which "the meaning of the construction is the combinatorial result of the basic senses and the syntactic relations of the lexical items within the construction" (Barkema, 1996: 138). For instance, the multiword unit *a black dog* is fully compositional since it can be understood by combining the basic senses of "black" and "dog" with the syntactic relationship between them.

- Pseudo-compositional constructions involve lexical items in which "the basic senses of the lexical items in the constructions play a role in, but form only parts of, their meanings" (Barkema, 1996: 139). For example, the multiword unit *bed and breakfast* refers to a specific type of accommodation that provides a systematic service, including but not limited to providing a bed and breakfast.
- Fully non-compositional constructions are those in which "the constructions do not have lexical items with basic senses that form (part of) their meanings" (Barkema, 1996: 139). An example of this type of construction is "put the cat among the pigeons".
- Partly non-compositional constructions contain at least one lexical item with a basic sense, but not all the basic and derived senses of the lexical items contribute to the overall meaning of the construction (Barkema, 1996: 140). For example, "a blind valley" is partly non-compositional since the basic sense of "blind" contributes to the overall meaning of the construction, while the basic sense of "valley" does not. This differs from pseudo-compositional constructions, where some of the basic or derived senses of the lexical items contribute to the overall meaning.

Compositionality plays a crucial role in the field of phraseology, as it enables researchers to differentiate between various types of phraseological units. Katz (1996) distinguished idioms from other phraseological units, as their meaning "is not a compositional function of the meanings of the idiom's elementary grammatical parts" (p. 275). However, the degree of compositionality may vary across different types of phraseological units. For example, Barkema's (1996) theory suggests that the compositionality of two adjectival modifiers, *black dog* and *red herring*, is entirely different, with the former being fully compositional, whereas the latter is fully non-compositional. While *black dog* denotes a dog whose color is black, *red herring* refers to a

misleading statement, question, or argument intended to divert a conversation from its original topic.

Svensson (2008) identified four defining dichotomies associated with non-compositionality, namely motivation/non-motivation, transparency/opacity, analyzability/unanalyzability, and literal/figurative meaning. The term “motivation” refers to the possibility of comprehending the meaning of each word in an expression once the meaning of the entire expression is known (Svensson, 2008:83). To describe an expression in which understanding the meaning of each word contributes to understanding the complete expression, Svensson suggested the term "motivatable." Conversely, an expression is unmotivated if it is impossible to comprehend the meaning of the expression by making sense of the meaning of each word included in it. For example, the expression *white wedding* is motivatable since the color white is commonly associated with purity and innocence.

Transparency denotes the attribute of an expression that enables a language user to understand it without any difficulties or previous knowledge, except for the comprehension of each separate word that constitutes the expression. In contrast, opacity describes the inability to reconstruct the meaning of an expression from the meanings of its composing elements. Similes, such as *as white as snow*, are typical examples of transparent expressions.

Analyzability is the characteristic of an expression where each individual element contributes to the meaning of the expression as a whole. The notion of "decomposition" is also used to describe analyzability. For instance, the expression *pop the question* is analyzable as each individual element in the expression contributes to the meaning of the expression as a whole. In this case, *pop* can be taken to mean to ask, and *question* can be taken to mean wedding proposal.

Another significant dichotomy of non-compositionality is literal and figurative meaning. Svensson argued that it is rather problematic to provide a definition to literal and figurative

meaning. Nevertheless, she believed that there is a relation between literal and compositional meaning as well as between figurative and non-compositional meaning. Gross (1996:11) argued that there are many sequences that a foreigner cannot interpret literally in any language, even if he or she knows the common meanings of all words that make them up. She stated that the “ordinary” meaning of the sentence “La moutarde lui monte au nez [‘The mustard goes up his nose’] does not allow one to conclude that the whole sentence means that a person is getting angry. Thus, we say that this sentence does not have a compositional meaning. Svensson then concluded that the feature ‘not having a literal interpretation’ must be equivalent to ‘having a figurative interpretation’.

Although the concept of compositionality has garnered considerable attention within scholarly investigations, the task of categorizing multiword units according to their varying degrees of compositionality remains complex. In this exposition, two prototypical challenges shall be expounded upon: ambiguity and context-dependence.

Determining the precise degree of compositionality for certain phraseological units can be difficult due to the potential ambiguity in their meanings. Some expressions may have multiple interpretations, making it challenging to assign a specific compositionality level accurately. Take the multiword unit *break a leg* for example. It is a commonly used idiom in English, typically used to wish someone good luck. Due to the ambiguity, two degrees of compositionality can be recognized: non-compositional and partially compositional. In the first interpretation, *break a leg* is treated as a non-compositional multiword unit with a high degree of idiomaticity. The literal meaning of breaking one's leg doesn't relate directly to the intended good luck wish. Therefore, this interpretation suggests that the phrase is largely opaque and not easily analyzable based on its individual components. In the second interpretation, *break a leg* is considered partially compositional. It acknowledges that the literal meaning of "break a leg" is unrelated to the

idiomatic usage but points out that the phrase still contains familiar and recognizable words. The degree of compositionality here might be seen as intermediate, as the phrase involves a well-known action ("break") and body part ("leg"), even though the specific idiomatic meaning isn't immediately obvious.

Determining the precise degree of compositionality for certain phraseological units can also be difficult due to the degree of compositionality of context dependent. Take the multiword unit *kick the bucket* for example. Depending on the context, two degrees of compositionality can be identified. If the phrase is used in a context where someone is actually kicking a physical bucket, the degree of compositionality is high. The expression is being used literally, and the individual components ("kick" and "bucket") contribute directly to the overall meaning of the action. In most other contexts, the phrase is used idiomatically to refer to someone's death. In this case, the degree of compositionality is low. The literal meaning of the individual components doesn't align with the intended idiomatic meaning, and the phrase becomes less transparent.

Evidently, the extent of compositionality undergoes influences from a multitude of factors, among which two (ambiguity and context dependency) have been elaborated upon earlier. This elucidation underscores the intricate nature of classifying multiword units according to their level of compositionality, revealing the array of challenges inherent in such categorization.

2.1.3 Restricted Collocability

Restricted collocability is a phraseological phenomenon where specific words tend to occur together within particular contexts or collocational patterns. These patterns demonstrate the predictable and regular ways in which words combine with one another in language use.

According to Barkema (1996:45), collocability can be defined as "the extent to which a lexical

item from an open class can be replaced in a construction with an alternative from the same class, such as a noun with another noun, a verb with another verb, and so on."

The collocability of multiword units varies according to the degree to which their component elements can be replaced by synonyms, near-synonyms, or antonyms. Barkema (1996) classified multiword units into three types based on their collocational openness: collocationally open, closed, and limited. For instance, the multiword unit *increase dramatically* is an example of a collocationally open unit where substituting the word *dramatically* with synonyms such as *significantly* or *excessively* does not significantly change the meaning. In contrast, the components of some units, like *red tape*, are collocationally closed. This means that substituting the word *tape* with a synonym like *ribbon* alters the entire construction's meaning and may only result in a literal interpretation. Finally, some units are collocationally limited, indicating that while their components can be alternated to some extent, there are limits to the available options. For instance, *generally speaking* can be replaced with *strictly speaking*, but alternatives like *elaborately speaking* are not typically used.

In summation, this section has undertaken an examination of the trio of pivotal criteria that delineate phraseology, namely polylexicality, compositionality, and restricted collocability. Remarkably, these three attributes collectively embody the essence of the phraseological domain. Importantly, the concept itself presents a multi-dimensional construct, and the attempt to classify phraseology singularly based on a sole criterion such as compositionality is fraught with a diverse range of challenges.

2.2 Categorization

This section provides an overview of the categorization schemes delineated by prior researchers within the realm of phraseology. Broadly, two primary methodological orientations

emerge: the phraseological approach and the distributional approach, each distinctively accentuating a facet intrinsic to the domain.

2.2.1 The phraseological approach

In the former Soviet Union, the study of phraseology was pioneered by scholars such as Vinogradov, who proposed a conceptual framework that characterizes phraseological units as a continuum ranging from highly rigid and impenetrable to flexible and transparent. This approach, commonly referred to as the "phraseological approach," utilizes linguistic criteria to distinguish various types of phraseological units. The phraseological approach of classifying multiword units involves focusing on their internal structure and composition. Additionally, multiword units are categorized based on their inherent semantic and syntactic properties, and their classification is guided by the principles of grammar and meaning.

Vinogradov (1953) classified phraseological units into phraseological fusions, phraseological unities, and phraseological combinations based on the relationship between the meaning of the unit and the meaning of its component parts. Similarly, Cowie (1981) categorized phraseological units and created the following continuum based on the degree of opacity and fixedness, with opacity and fixedness increasing along the spectrum:

- Free combinations (e.g., *drink tea*): The restriction on substitution can be specified on semantic grounds and all component elements of the word combination are used in a literal sense.
- Restricted collocations (e.g., *perform a task*): It enables certain with arbitrary limitations; At least one element has a non-literal meaning, and at least one element is used in its literal sense.

- Figurative idioms (e.g., *do a U-turn*): Mere substitution of the elements is pertained; The combination has a figurative meaning, but preserves a current literal interpretation.
- Pure idioms (e.g., *blow the gaff*): No substitution of the elements is pertained; The combination has a figurative meaning and does not preserve a current literal interpretation.

Due to the lack of a consensus on the terminology used in phraseology, researchers have encountered difficulties in defining and categorizing multiword units. Different typologies have been developed for diverse purposes, including lexicographic, pedagogical, and psycholinguistic. Regarding the categorization from the perspective of the phraseological approach, Cowie (1998, 2001) proposed a typology of word combinations, which included composites and formulae. Composites were divided into restricted collocations, figurative idioms, and pure idioms, with decreasing degrees of transparency and fixedness. Restricted collocations were defined by their limited collocability, such as adjective-noun word combinations (e.g., *dramatic increase*). Figurative idioms had figurative meanings and were characterized by non-transparency and a certain degree of fixedness, such as *have a seat*. Pure idioms, such as "the early bird gets the worm", were the least transparent and non-compositional, meaning that the elements in the idioms could not be replaced with other elements. Cowie further categorized formulae into routine formulae, which were expressions tied to recurrent social situations (e.g., *good morning*), and speech formulae, which were expressions used to organize discourse and indicate speakers' attitudes (e.g., *I beg your pardon*). A diagram illustrating Cowie's typology is presented in Figure 1.

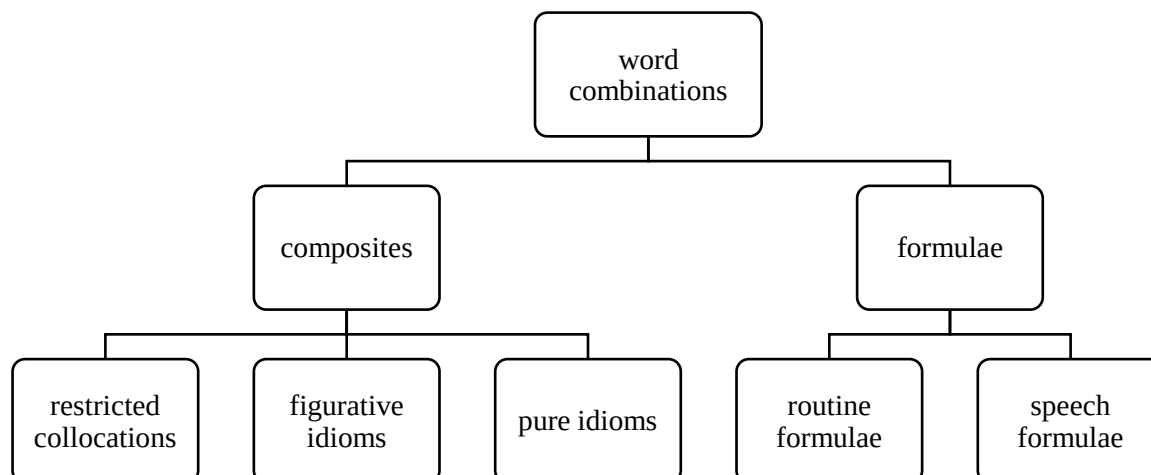


Figure 2.1 Cowie's (1998, 2001) typology of phraseological units

Mel'čuk (1998) introduced an influential taxonomy that distinguished set phrases or phrasemes into two categories: semantic phrasemes and pragmatic phrasemes or pragmatemes. According to Mel'čuk, semantic phrasemes refer to set phrases whose meaning is freely chosen but the expression for that meaning is not. This category comprises full phrasemes or idioms, quasi-phrasemes or quasi-idioms, and semi-phrases or collocations. Full phrasemes contain semantic phrasemes whose signified does not include the signified of their components in a semantically dominant position. For instance, the signified of *spill* and *bean* are not included in the signified of *spill the bean*. Quasi-phrasemes include the signified of their components plus an additional signified. An example of a quasi-phraseme or quasi-idiom is "bed and breakfast," which includes not only the signified of its components, but also additional signified such as heat, electricity, and the service of hotel waiters, among others. Semi-phrasemes or collocations are semantic phrasemes in which the global signified is derived from the signified of one of their components A and a signified 'X,' with the other component B only expressing 'X' contingent on A. For instance, the collocation heavy rain is constructed out of the signified of rain and a signified X that the adjective heavy only expresses contingent to the noun rain. In contrast,

pragmatic phrasemes or pragmatemes are context-dependent and their meaning corresponds to Cowie's formulae. Thus, they are "non-compositional pragmatically" (Mel'čuk, 1998:29). For instance, *good morning* is an example of a pragmatic phraseme.

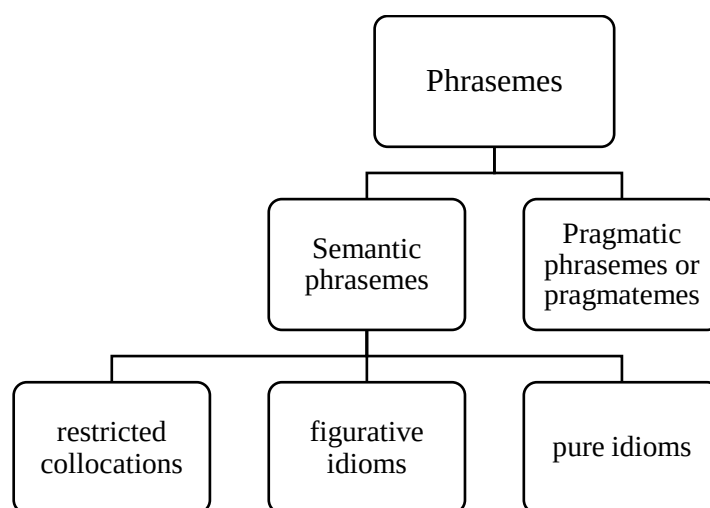


Figure 2.2 Mel'čuk's (1998) typology

2.2.2 The phraseological approach

In contrast to the phraseological approach of categorizing multiword units, a more recent approach, known as the "distributional approach" or "frequency-based approach," has emerged. This approach uses a bottom-up, corpus-driven methodology to examine phraseological units. Rather than pre-categorizing different types of units, this approach examines the lexical co-occurrence of words in phrases. This approach, as demonstrated by Altenberg (1998), expands the possibilities for phraseology research, including units that were previously outside the domain of phraseology in the traditional approach. Altenberg analyzed recurrent word-combinations in the London-Lund Corpus of Spoken English by extracting all n -word ($n \geq 2$) word-combinations occurring more than once in identical form and focusing specifically on 3-word combinations that occurred at least ten times in the corpus. The distributional approach, which examines the usage

patterns of phraseological units, and as illustrated in Figure 3, that has gained considerable recognition in the field.

In this categorization, two distinct categories are defined, namely n-gram or cluster analysis, and co-occurrence analysis. The n-gram or cluster analysis focuses on studying

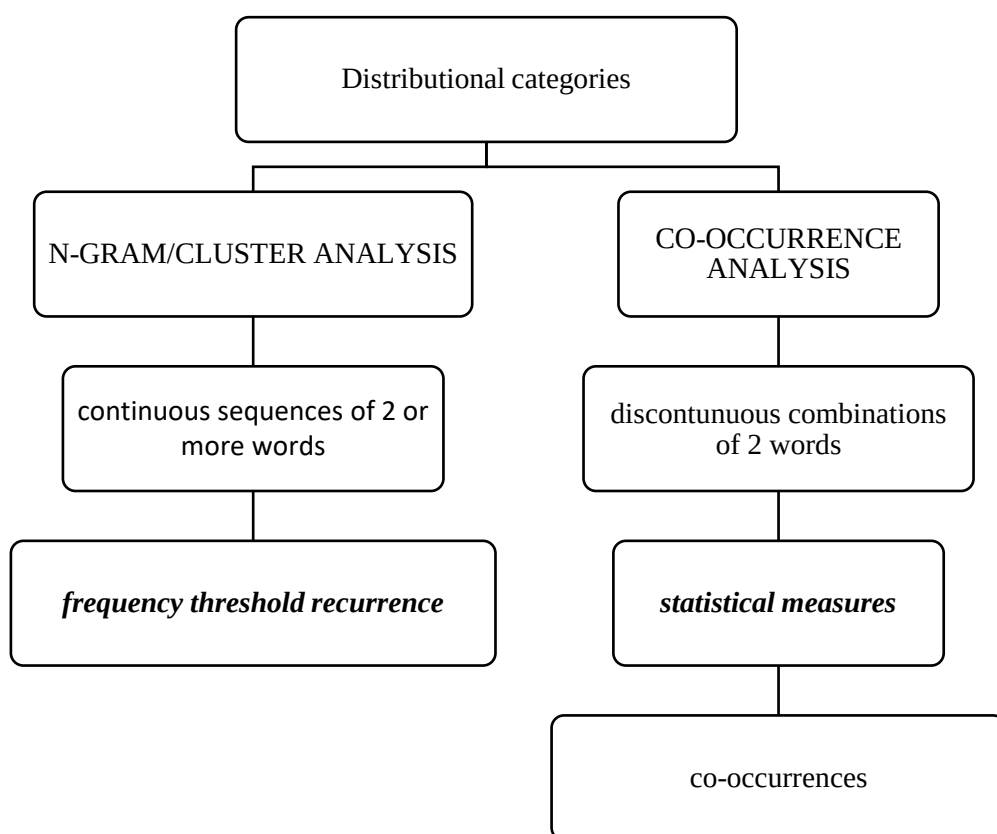


Figure 2.3 Distributional categories

continuous or adjacent sequences of two or more words. These extracted sequences are defined as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (Biber et al., 1999:990). The researcher determines the length of the sequences, frequency threshold, and other relevant parameters depending on the research objectives. For example, Simpson-Vlach and Ellis (2010) set the length of the sequences to 3, 4, and 5 words in the Michigan Corpus of Academic Spoken English, plus selected British National Corpus files that

occurred at least 10 times per million words in the target corpus, to create a list of exclusively academic word sequences.

In contrast, co-occurrence analysis focuses on studying non-adjacent combinations of two words using statistical measures such as mutual information (Rogers et al., 2021) or log likelihood ratio (Durrant, 2009), which are referred to as 'collocations' or 'collocates.' For instance, Durrant (2009) created an academic collocation list that takes into account variations across disciplines. He followed Jones and Sinclair's (1974) widely used precedent of limiting 'co-occurrence' to occurrences within a four-word span and required any potential collocation included in the final list to have a mutual information score of at least 4.

To sum up, the study of phraseology has evolved over time, from the phraseological approach that classified phraseological units based on their rigidity and opacity, to the distributional approach that uses a corpus-driven methodology to examine lexical co-occurrence in phrases. Different typologies have been proposed, each with its own set of categories and criteria, which highlights the difficulties in defining and categorizing phraseological units.

2.3 Zooming in on Collocations

This subchapter specifically narrows the focus to collocations, which constitute the primary area of interest in this research. Building upon the previously discussed phraseological approach and distributional approach to studying phraseology, the definitions of "collocation" presented in this chapter align with those introduced earlier.

2.3.1 The phraseological approach

This section discusses the approach of investigating phraseology that were strongly influenced by the classical Russian theory (e.g., Vinogradov, 1953) and were referred to as “phraseological approach”. Researchers adopting this approach (e.g., Cowie, Mel’čuk) deemed

that collocations constitute a continuum between combinations that combine freely (i.e., free combinations) to combinations that combine highly restricted (i.e., idioms). However, they differ on how the continuum is exactly defined and operationalized.

2.3.1.1 Definitions of ‘collocation’

Some typical definitions of ‘collocation’ adopting such approach are shown in Table 2.1.

Proposer	Definitions
Van Roey (1990:48)	Collocation is the term we shall now use both for the linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than that of its 'synonyms' because of constraints which are not on the level of syntax or conceptual meaning but on that of usage, as also for the word combinations which represent this phenomenon. Collocational meaning then is the meaning of a word as far as it is determined by the lexical company it keeps.
Gramley and Patzold (1992:61)	Collocation refers to combinations of two lexical items which make an isolable semantic contribution, belong to different word classes and show restricted range.
Cowie (1994:369)	Collocations are associations of two or more lexemes (or roots) recognized in and defined by their occurrence in a specific range of grammatical construction. HEAV- + RAIN is one such abstract composite, realized in the patterns <i>heavy rain</i> and <i>rain heavily</i> (Mitchell, 1971). Though transparent and (usually) lexically variable (c.f. <i>light rain</i> , <i>a heavy shower</i>), they are characterized by arbitrary limitation of choice at one or more points, as in <i>light exercise</i> / <i>heavy exercise</i> (Cowie, 1981).
Howarth (1996:47)	Combinations in which one component is used in its literal meaning, while the other is used in a specialized sense. The specialized meaning of one element can be figurative, delexical or in some way technical and is an important determinant of limited collocability at the other. These combinations are, however, fully motivated.
Mel’čuk (1998:30)	A collocation AB of language L is a semantic phraseme of L such that its signified ‘X’ is constructed out of the signified of one of its two constituent lexemes – say, of A – and a signified ‘C’ [$X = A \oplus C$] such that the lexeme B expresses ‘C’ only contingent on A .

Table 2.1 Definitions of ‘collocation’ proposed with the phraseological approach.

Drawing upon the aforementioned exemplar definitions of 'collocation' within the framework of the phraseological approach, several patterns emerge. Firstly, it becomes evident that a collocation invariably constitutes a confluence of two or more words. Secondly,

collocations are combinations of words that function below the level of the sentence. Thirdly, the elements of a collocation tend to appear together and forms a relatively stable combined unit which is not usually separable. Fourthly, regarding the meaning of the collocation and the meaning of the elements of the collocation, the meaning of the collocation is not simply the addition of its elements' meaning.

Furthermore, Cowie (1994) introduced a classification of collocations based on specific grammatical relationships. Instances such as *light rain* and *heavy shower* exemplify collocations falling under the adjective + noun grammatical relationship. However, Cowie observed that while these grammatical relationships may appear somewhat arbitrary, they are simultaneously bounded by constraints.

Moreover, certain scholars have established criteria governing the elements of a collocation. Notably, Gramley and Patzold (1992) as well as Cowie (1994) stipulated that the elements of a collocation must exclusively comprise lexical items. This requirement entails the exclusion of grammatical words from collocational pairs. For instance, according to Gramley and Patzold (1992), *big apple* qualifies as a collocation given that both *big* and *apple* are lexical words. In contrast, the phrase *at present* does not meet the criteria for collocation status, as *at* serves as a functional word rather than a lexical item.

2.3.1.2 Classification of collocations

As illustrated above, linguists proposed different definitions of 'collocation' within the phraseological approach. Depending on their definitions, they have proposed classifications of collocations, which can be generally classified into two main categories: syntactic classifications and semantic classifications.

2.3.2.2.1 Syntactic classifications

Benson et al (1986) is one of the most typical to classify collocations according to their syntactic patterns. According to them, collocations were generally classified into two main categories: grammatical collocations and lexical collocations. Grammatical collocations refer to “a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or a clause” (Benson et al, 1986:1). Lexical collocations, in contrast to grammatical collocations, normally do not contain prepositions, infinitives, or clauses. Typical lexical collocations consist of nouns, adjectives, adverbs, and verbs. Table 2.2 shows in detail how different types of grammatical collocations are comprised of and corresponding examples.

	Formulation	Example
G1	noun + preposition	<i>blockade against</i>
G2	noun + <i>to</i> infinitive	<i>...a pleasure to see you.</i>
G3	noun + <i>that</i> clause	<i>...read a book that I like.</i>
G4	preposition + noun	<i>by accident</i>
G5	adjective + preposition	<i>They are angry at me.</i>
G6	adjective + <i>to</i> infinitive	<i>It is good to see you.</i>
G7	adjective + <i>that</i> clause	<i>I'm afraid that...</i>
G8	collocations consist of 19 English verb patterns	<i>He sent him it.</i>

Table 2.2 Grammatical collocations proposed by Benson et al (1986)

Table 2.3 shows how different types of lexical collocations are formulated and corresponding examples.

	Formulation	Example
L1	verb + noun/pronoun (or <i>prep</i> phrase)	<i>fly a kite</i>
L2	verb meaning essentially eradication and /or nullification + noun	<i>reverse a law</i>
L3	adjective + noun	<i>good book</i>
L4	noun + verb	<i>blood circulates</i>
L5	indicate the unit that is associated with the noun	<i>a pack of dogs</i>
L6	adverb + adjective	<i>deeply absorbed</i>
L7	verb + adverb	<i>apologize humbly</i>

Table 2.3 Lexical collocations proposed by Benson et al (1986)

The approach proposed by Benson et al (1986) of classifying collocations into grammatical collocations and lexical collocations is defined by the syntactic patterns of elements that compose the collocation. Grammatical collocations are comprised of one content word (i.e., words that carry significant meaning in a sentence. In this case, a noun or adjective) and another grammatical word (i.e., words that primarily serve to convey grammatical relationships and syntactic structure within a sentence. In this case, a preposition) or a grammatical structure such as an infinitive or a clause. Lexical collocations, on the other hand, are comprised of two content words. No grammatical words are involved in the process of formulating lexical collocations.

2.3.2.2.2 Semantic classifications

Apart from classifying collocations depending on the syntactic patterns of the collocations' elements, researchers also proposed approaches to classify collocations depending on the semantic meanings of the collocations' elements.

Mel'čuk (1998) is a typical example of classifying collocations semantically. He defined collocations "a collocation **AB** of language **L** is a semantic phraseme of **L** such that its signified 'X' is constructed out of the signified of one of its two constituent lexemes – say, of **A** – and a signified 'C' [$X = A \oplus C$] such that the lexeme **B** expresses 'C' only contingent on **A**" (Mel'čuk, 1998:30). According to his definition on collocations, it is worth noting that the meaning of the collocation **AB** depends on the two constituents **A** and **B**. However, **B** expresses **C** only contingent on **A**. The formulation 'B expresses C only contingent on A' covers for major cases, which correspond to the following four major types of collocations:

1. Either $C \neq B$, i.e. **B** does not have (in the dictionary) the corresponding signified.

And [(a) **C** is empty, that is, the lexeme **B** is, so to speak, a semi-auxiliary selected by **A** to support it in a particular syntactic configuration;

Or (b) **C** is not empty but the lexeme **B** expresses **C** only in combination with **A** (or with a few other similar lexemes)]

2. Or **C** = **B**, i.e. **B** has (in the dictionary) the corresponding signified.

AND [(a) **B** cannot be expressed with **A** by any otherwise possible synonym of **B**;

OR (b) **B** includes (an important part of) the signified A, that is, it is utterly specific, and thus **B** is ‘bound’ by **A**].

Examples of collocations that fall into 1(a) category includes *take a step*, *give a look*. In these two examples, the meanings of the verbs *take* and *give* are weakened and they both lose their original meanings. They become what other linguists call *light verb constructions*. Example of collocations that fall into 1(b) category includes *black coffee* and *red tape*. In these examples, the collocator develops a specialized meaning only when combining with the base. Examples of collocations that fall into 2(a) category includes *strong coffee*. In this example, the collocator keeps its meaning on the dictionary but cannot be replaced by any synonyms in the way that replacing *strong* with *powerful* in *strong coffee* does not make sense. Examples of collocations that fall into 2(b) category includes *horse neighs*, *aquiline nose*, *rancid butter*, or *artesian well*. In these examples, the collocator is bound to the base by including the base in its meaning.

2.3.2 The distributional approach

Halliday (1961:276)	Collocation is the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur, at n removes (a distance of n lexical items) from an item x, the items a, b, c.
Jones and Sinclair (1974:19)	‘Collocation’ is the co-occurrence of two items in a text within a specific environment. “Significant” collocation is regular collocation between items, such that they co-occur more often than their respective frequencies and the length of text in which they appear would predict.
Hoey (1991:6-7)	The relationship a lexical item has with items that appear with greater than random probability in its (textual) context.
Sinclair (1991:170)	Collocation is the occurrence of two or more words within a short space of each other in text. The usual measure of proximity is a

	maximum of four words intervening. Collocation can be dramatic and interesting because unexpected, or they can be important in the lexical structure of the language because of being frequently repeated.
--	--

Table 2.4 Definitions of ‘collocation’ proposed with the phraseological approach.

It is obvious from Table 2.4 that among definitions proposed by linguists adopting the distributional approach, they all address one point that is not present in the phraseological approach, namely, the frequency of occurrence. Instead of defining collocations according to their elements both syntactically and semantically as used by the phraseological approach or traditional approach, the distributional approach highlights the high-frequency co-occurrence of two or multiple words. That is why the distributional approach is often referred to as frequency approach. The distributional approach gains more and more attention and has been adopted as the primary definition of various studies towards collocations (e.g., Shin & Nation, 2008; Durrant, 2009; Ackermann and Chen, 2013; Chon & Shin, 2013).

Over the years, different researchers adopted the distributional approach of defining collocations and contributed to the study of collocations in various ways, especially in creating high-frequency academic collocation lists. Shin and Nation (2008:341) defined collocations as “a group of two or more adjacent words that occur frequently together”. A collocation is operationalized to be made of two parts – a pivot word which is the focal word in the collocation and its collocate(s), the word or words accompanying the pivot word. Similarly, Durrant (2009:161) defined collocations as “word pairs that co-occur at least moderate frequency”. They were operationalized as word pairs which co-occurred within a four-word span of each other. Ackermann & Chen (2013:237) defined collocations as “a single word (node word) that tends to co-occur in the span of ± 3 words from the reference word”. Finally, Chon & Shin (2013:17) defined collocations as “a group of two or more adjacent words that occur frequently together”.

They were operationalized to be made of two parts – a pivot word which is the focal word in the collocation and its collocate(s), the word or words accompanying the pivot word.

2.3.3 Definition of ‘collocation’ in this research

In this research, a distributional definition of ‘collocation’ is drawn from Paquot (2019:126-127) that ‘collocation’ refers to “word combinations used in three grammatical relations, i.e., adjectival modifiers (*amod*: adjective + noun), adverbial modifiers (*advmod*: adverb + adjective, adverb or verb) and verb + direct object (*dobj*)”. The *amod*, *advmod*, and *dobj* Stanford typed dependency relations are binary grammatical relations between a governor and a dependent (see De Marneffe and Manning, 2013). Examples of collocations adopting this definition is shown in Table 2.5.

	<i>amod</i>	<i>advmod</i>	<i>dobj</i>
Examples	academic research, academic journal	increase dramatically, systematically influence	have access, reach goal

Table 2.5 Example of collocations in this research

In summary, Chapter 3 reviews previous literature concerning phraseology and collocation. Three specific groups of literature were reviewed, namely defining criteria of phraseology, categorizations of phraseological units, and definitions and categorizations of collocations. Three defining criteria were reviewed in the first section, namely polylexicity, compositionality, and restricted collocability. Two approaches of categorizing phraseological units were reviewed in the second section, namely phraseological approach and distributional approach. The last section zooms the perspective into collocations, the main focus of this research. Typically, two approaches of studying collocations are reviewed, i.e., phraseological approach and distributional approach. The phraseological approach focuses on analyzing the elements of collocations syntactically and semantically. The distributional approach, on the other hand, focuses on the co-occurrence of two lexical items. Within the domain of these two

approaches, a substantial number of contributions were made to the study of collocations. The next chapter delves into the study of multiword units with second language, including how learners use different types of collocations.

3 Phraseology and Complexity in L2 Language

Phraseology, the study of multiword units that exhibit recurrent patterns of co-occurrence and exhibit a level of stability beyond the sum of their individual components, is a pivotal area in linguistics that sheds light on the intricate nature of learner language usage. In the realm of learner corpus research (LCR), investigating how learners employ multiword units, such as collocations and lexical bundles, provides a nuanced lens through which we can understand the development and proficiency of learners' linguistic competence.

In recent years, the utilization of learner corpora has emerged as a powerful methodology to gain insights into the linguistic behavior and performance of language learners. These specialized corpora offer a unique window into the challenges and successes encountered during the process of language acquisition. One compelling avenue of research within learner corpus studies is the examination of how learners deploy multiword units in their writing and speech.

This chapter delves into the pivotal intersection of phraseology and L2 language. Section 3.1 introduces how previous researchers investigated the use of various multiword units by EFL learners. Section 3.2 zooms into phraseological complexity – an innovative intersubject of phraseology and L2 complexity research.

3.1 Studies of Multiword Units in Learner Language

The exploration of phraseology and its influence on language learning has garnered significant interest within the domain of LCR in recent times. As a result, a plethora of studies have delved into the role of phraseology. Additionally, researchers have extensively employed collocations and learner corpora in the field of LCR to examine how learners utilize various types of collocations and the valuable insights they can provide regarding LCR.

Paquot and Granger (2012) identified two main types of phraseological patterns that have been studied in the literature on phraseological units and learner corpus research: co-occurrence and recurrence. The co-occurrence pattern involves the simultaneous occurrence of (typically) two lexical items, which may or may not be adjacent to each other. Studies adopting the co-occurrence pattern include both the traditional phraseological approach, where collocations or restricted collocations are defined as lexically constrained combinations allowing limited substitution within a specific grammatical construction (e.g., *perform a task* is acceptable whereas *make a task* is not), and the distributional approach, which relies on a frequency-based perspective (e.g., Sinclair, 1987). Research utilizing the co-occurrence pattern has demonstrated that learners' use of co-occurring combinations is characterized by a combination of underuse, overuse, and misuse (Paquot and Granger, 2012:136). In other words, learners may fail to employ certain collocations adequately, overuse them inappropriately, or misuse them in non-standard ways.

The recurrence pattern, in contrast, is defined as the repetition of contiguous strings of words of a specific length, such as bigrams or trigrams (Paquot & Granger, 2012:138). This concept aligns with what Biber et al. (1999) referred to as lexical bundles. The recurrence pattern entails extracting adjacent sequences of words of a given length, which can include both grammatically complete strings (e.g., *on the one hand*) and incomplete strings (e.g., *it should be noted that*). Paquot and Granger noted that studies adopting the recurrence pattern can be challenging to compare, as they involve extracting lexical bundles of varying lengths, and the specific settings used for extracting these bundles may vary considerably. Therefore, the analysis of recurrence patterns requires careful consideration and standardization to ensure consistency across different studies.

Furthermore, scholarly inquiries have delved into the examination of multiword unit utilization by learners, with the aim of discovering discernible patterns underlying their usage. These investigations have yielded a spectrum of intriguing patterns. This section scrutinizes three of these patterns, namely the overutilization of selected high-frequency multiword units, the underutilization of selected low-frequency but high-association multiword units, and the influence exerted by learners' respective first languages (L1s).

One group of research resulted that EFL learners tend to overuse certain high-frequency multiword units. One of the pioneering studies examining collocations in learner language was conducted by Granger (1998). The author focused specifically on the utilization of amplifier collocations, which refer to adverbs ending with "-ly" that function as modifiers (e.g., *perfectly natural*). The study compared the production of these collocations by native speakers of English and French EFL learners within the framework of Contrastive Interlanguage Analysis (CIA; Granger, 1996), which is a linguistic approach used in second language acquisition (SLA) research to examine the differences and similarities between the interlanguage of second language learners and both their native language and the target language they are learning. To carry out the investigation, Granger employed the Louvain Essay Corpus as the native English corpus and a subcorpus of the International Corpus of Learner English as the learner English corpus. Each corpus comprised approximately 250,000 tokens. The retrieval software TACT was utilized to extract and analyze the individual modifiers. The results revealed that English native speakers exhibited a greater variety of amplifier usage, while French EFL learners demonstrated more limited usage by employing a smaller range of amplifiers and repetitively relying on specific ones such as *completely* and *highly*. This suggests that there are discernible differences in collocation usage between EFL learners and native speakers of English, indicating that collocations can serve as a valuable starting point for analyzing learner language.

Lee (2006) examined the use of amplifier collocations in English produced by English native speakers and Korean EFL learners. The result shows that compared to English native speakers, Korean EFL learners make use of a limited number of high-frequency amplifiers in their writing. In another study, Chen and Baker (2010) examined the usage of lexical bundles in academic writing. The lexical bundles were operationalized as adjacent word combinations of several words, which are usually 2 to 6 words. They extracted 4-grams (i.e., adjacent word combinations of four words) with a frequency of 25 per million tokens or higher. The results showed that, in contrast to native speakers who employed various types of collocations, EFL learners tended to use a narrower range of collocations that align with the researchers' interests.

Another group of research resulted that EFL learners tend to underuse certain low-frequency but high-association multiword units. In addition to examining collocations based solely on frequency, researchers have incorporated another statistical measure into their analysis: association. Durrant and Schmitt (2009:167) incorporated the concept of association between collocations' elements, defined as "the likelihood of collocations appearing together with a probability greater than chance". They focused on adjacent adjective + noun and noun + noun collocations produced by English native speakers and EFL learners, considering two statistical constructs: frequency and association, operationalized as the relative frequency and mutual information (MI) score plus t-score respectively. The findings revealed that EFL learners heavily relied on high-frequency collocations but underused less frequent yet strongly associated collocations, which are likely to be highly salient for native speakers.

Granger and Bestgen (2014) expanded the types of collocations to be analyzed and investigated adjacent adjective + noun, noun + noun, and adverb + adjective collocations produced by English native speakers and EFL learners, focusing on two dimensions: frequency and association, operationalized as relative frequency, MI score, and t-score. Similarly, Bestgen

and Granger (2018) extended this research by analyzing the same collocation types produced by the same group of EFL learners but at different time periods, employing a longitudinal approach. The findings indicated that texts produced by learners at higher proficiency levels tended to contain fewer high-scoring t-score collocations but more high-scoring MI collocations.

Another group of research resulted that learners' L1 background has an impact on their use of multiword units. Parkinson (2015) investigated the utilization of noun-noun collocations (e.g., *family members*) by English native speakers and EFL learners whose first languages were Mandarin, Spanish, or Tswana. The findings indicated that learners' first languages had an influence on the production of collocations, with learners whose first language allowed for noun-noun phrases producing significantly more of them in English compared to those whose first language did not.

3.2 Studies of Phraseological Complexity

Indeed, a substantial body of research within the field of learner corpus research has concentrated on examining the usage of collocations produced by EFL learners from diverse backgrounds, including variations in proficiency levels and first languages. These studies have recognized that phraseological units, such as collocations, have emerged as crucial indicators of EFL learners' linguistic proficiency. Despite that, phraseological competence has not received much attention from most language testers, as noted by Paquot (2018:29), who stated that "its development has not received the attention it deserves in the CEFR". To address this issue, Paquot (2019) proposed the concept of phraseological complexity, which connects L2 phraseology research with L2 complexity research. This construct, based on Ortega's (2003:124) definition, measures "the range of phraseological units that surface in language production and the degree of sophistication of such phraseological units", and was operationalized as

phraseological diversity and phraseological sophistication. The definition of "phraseological units" used in this research was adopted from Gries (2008:6) as "the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance". Paquot conducted a study using a corpus of linguistics term papers written by French L2 EFL learners to explore phraseological complexity and evaluated three types of phraseological units that have been found to be challenging for L2 learners: adjectival modifiers (adjective + noun, e.g., *clear evidence*), adverbial modifiers (adverb + adjective, verb, or adverb, e.g., *statistically significant*), and direct objects (verb + noun, e.g., *make efforts*). These three syntactic relationships are abbreviated as *amod*, *advmod*, and *dobj*, respectively, according to Stanford typed dependency (De Marneffe & Manning, 2013).

Phraseological diversity was used to measure the depth of learners' phraseological competence. Specifically, it represents the number of unique phraseological units to the total number of phraseological units, indicating the variety of different phraseological units used by learners. It was operationalized as the root type-token ratio of the three types of syntactic relations, which represented the number of unique phraseological units to the total number of phraseological units, indicating the variety of different phraseological units used by learners.

Phraseological sophistication was used to measure the breadth of learners' phraseological competence and referred to the ability of learners to use advanced and appropriate phraseological units that are specific to the topic and style of writing, rather than general everyday vocabulary. The study employed two methods to measure phraseological sophistication. The first method involved defining sophisticated phraseological units as academic collocations and operationalized as word combinations that appear in Ankermann and Chen's (2013) Academic Collocation List (ACL). The ACL includes 2,468 highly frequent and pedagogically relevant lexical collocations

extracted from academic writing and was evaluated using a corpus-driven and expert-judged approach. Phraseological sophistication was measured as ratios of the number of sophisticated *amod*, *advmod*, and *doj* tokens to the total number of *amod*, *advmod*, and *doj*, both on a token-based and type-based level. The study also employed pointwise Mutual Information (PMI) for *amod*, *advmod*, and *doj* dependencies, a measure borrowed from information theory that compares the probability of observing word pairs together with the probabilities of observing them independently. Three mean MI scores were calculated for each learner text based on all the different word pairs found in the *amod*, *advmod*, and *doj* dependencies.

The result of phraseological diversity indicated that there was no significant difference in phraseological diversity across proficiency levels. The result of phraseological sophistication indicated that the PMI-involved phraseological sophistication exhibited a significant increase across proficiency levels. The ACL-involved phraseological sophistication exhibited a systematic ratio increase across proficiency levels of all dependencies. This demonstrates that as English EFL learners progress to higher proficiency levels, there is a systematic increase in the proportion of ‘sophisticated collocations’ (i.e., collocations included in the ACL), both type- and token-based, suggesting that phraseological sophistication is an effective measure of learners’ phraseological competence and can be used to evaluate their ability to use appropriate phraseological units in academic writing. Nevertheless, no statistical significance was found in any intergroup ratio increase.

For the statistical insignificance amid systematic ratio increase resulted from the ACL-involved phraseological sophistication, Paquot believed that the small size of the ACL along with the small coverage of the ACL in the source corpus is one of the causes. The ACL contains 2,496 collocations, and covers 1.4 % of total tokens in the source corpus. Compared to the coverage of

other lists used in examining learner writing sophistication (e.g., The Academic Word List, which covers approximately 10% of academic texts), this coverage rate is relatively small.

In addition to utilizing phraseological complexity to describe L2 performance at different proficiency levels, Paquot (2019) also compared phraseological complexity with traditional measures of syntactic and lexical complexity. The outcomes of the analysis revealed that while no significant difference in lexical or syntactic complexity across proficiency levels was observed, a marked increase in PMI-involved phraseological sophistication was observed as proficiency levels progressed, suggesting that the construct of phraseological complexity can be a useful index of L2 proficiency at the upper levels of proficiency.

Paquot (2019) has opened new avenues for future research on phraseological complexity in L2 settings, thereby creating endless possibilities for investigating L2 phraseology and complexity. Building on this foundation, Paquot et al. (2020) conducted a longitudinal study on phraseological complexity, with a specific focus on direct object constructions, using the Longitudinal Database of Learner English (LONGDALE). The research investigated the impact of proficiency and time spent on learning English on the phraseological sophistication of verb + object relations in the writing of French EFL learners. The phraseological complexity was measured using mutual information at three different data collection points. It is worth noting that no ACL-derived phraseological sophistication method was used. The study employed a mixed-effects model to control for topic, time, and proficiency, and the results indicated that learner proficiency was a better predictor of phraseological complexity than the actual data collection point at which the essay was written.

Although the majority of the research on phraseological complexity discussed earlier centered on operationalizing it in relation to the written English of English EFL learners, there have been several attempts to expand the scope of phraseological complexity in a broader sense,

starting with examining different languages and with the objective of seeing whether the phraseological complexity measures which were originally developed by Paquot (2018, 2019) for L2 English, would also be predictive of proficiency in other languages. Vandeweerd et al (2021) conducted replication research of Paquot (2019) in a corpus of L2 French argumentative essays, in which phraseological complexity was operationalized as the diversity (root type-token ratio; RTTR) and sophistication (pointwise mutual information; PMI) of three types of grammatical dependencies: adjectival modifiers, adverbial modifiers and direct objects. Utilizing a L2 French corpus, the Leerdercorpus Frans, Vandeweerd operationalized the two approaches of phraseological complexity and found that significant increase was observed in the mean PMI of direct objects and the RTTR of adjectival modifiers across proficiency levels. In addition, phraseological complexity's relationship between other traditional complexity measures, i.e., morphological, lexical, and syntactic complexity, were investigated. Contrary to Paquot (2019) which resulted that phraseological measures are better predictors of proficiency than traditional complexity measures, this research found that the most important predictors of learners' L2 French performance also included traditional measures of complexity. It is worth noting that in this phraseological complexity replication study conducted on L2 French, no measure of phraseological sophistication based on a list of sophisticated academic collocations was used.

Similarly, Rubin et al (2021) conducted replication study of Paquot (2019) in a corpus of L2 Dutch, revealing that the measures of L2 Dutch phraseological complexity contribute substantially to the prediction of Dutch learners' proficiency assessing regression model, suggesting that "complexity measures tapping into phraseological phenomena will help to better model the full range of learner proficiency" (p. 120). It is also worth noting that in this replication study conducted on L2 Dutch, no measure of phraseological sophistication based on a list of sophisticated academic collocations was used.

Despite the significant role of phraseology in L2 complexity research, previous studies were mostly cross-sectional and focused only on the written mode. To address these limitations, Vandeweerd et al (2022) conducted a longitudinal and multitask-based research of phraseological complexity on L2 French. The study involved recruiting a specific cohort of L2 French learners who were instructed to complete a task consisting of a written argumentative essay, a semi-guided oral interview, and a picture-based oral narrative at three distinct time points over a duration of 21 months. Syntactic co-occurrences of adjectival modifiers and direct objects were extracted to analyze phraseological complexity. Results indicated that phraseological complexity performed differently across oral and written tasks, and no significant increase in phraseological diversity was observed over the 21-month period. There was a slight increase in phraseological sophistication, but only for direct objects. These findings emphasized the importance of task characteristics in measuring phraseological complexity.

To date, the only investigation of phraseological sophistication through the lens of an academic collocation list stems from Paquot (2019). However, subsequent explorations into the domain of phraseological sophistication within the contexts of L2 French and L2 Dutch have not used the academic collocation list-based methodology introduced by Paquot (2019). This departure can be attributed to the absence of dedicated lists for this analytical framework. It is noteworthy that the Academic Collocation List (Ackermann & Chen, 2013) employed in Paquot's study was designed not for research purposes, but rather as pedagogical resources catering to the needs of English for Academic Purposes (EAP) students and educators. Consequently, a research gap emerges—a lack of a purpose-built academic collocation list catering to research needs, particularly in its application to approximating L2 EFL writings in studies investigating phraseological sophistication.

In conclusion, this chapter discusses the interaction between phraseology, learner corpus research, and L2 complexity research. Extensive research has scrutinized the utilization of multiword units among the writings of EFL learners from diverse linguistic backgrounds. These investigations have yielded a plethora of findings, with this chapter delving into three specific outcomes: the overutilization of selected high-frequency multiword units, the underutilization of selected low-frequency but high-association multiword units, and the influence exerted by learners' respective L1s.

Additionally, phraseological complexity has demonstrated to be an effective indicator of learners' linguistic competence, presenting a promising avenue for future investigations in the domain of phraseology and L2 complexity. Nevertheless, a notable trend emerges from previous investigations into phraseological sophistication within the context of L2 French or Dutch. It becomes evident that none of these studies employed measures derived from an academic collocation list in French or Dutch. This tendency can likely be attributed to the absence of such lists in French and Dutch. Interestingly, in the case of English, Paquot (2019) stands as the solitary instance where the Academic Collocation List was used to approximate phraseological sophistication. However, it is essential to note that the intended purpose of the ACL was not for examining L2 writings; rather, its inception was oriented toward pedagogical utility, catering to EAP learners and educators.

Consequently, a research gap emerges—a compelling need to build an academic collocation list designed to examine L2 writings by EFL learners. Such a list promises to bridge an existing gap, equipping linguists with a tailored instrument to explore the intricacies of phraseological sophistication within English L2 writing assessments.

4. Multiword Unit Lists Construction

This chapter is devoted to reviewing previous literature concerning multiword unit lists constructions. Section 4.1 reviews the various multiword unit lists built by previous researchers and section 4.2 zooms in on the detailed statistical constructs and their corresponding operationalizations in previous research.

4.1 Multiword unit lists constructed in previous research

Shin and Nation (2008) is one of the earliest attempting building a multiword unit list, aiming at building a list of the highest frequency collocations of spoken English. Defining ‘collocation’ as “a group of two or more words that occur frequently together, and it is not restricted to two or three word sequences” (p. 341), they further operationalized collocations to be made of two parts – a pivot word which is the focal word in the collocation and its collocate(s), the word or words accompanying the pivot word. The pivot words were selected as the most frequent 1,000 content words according to Leech, Rayson, and Wilson (2001) and were set as a noun, a verb, an adjective, or an adverb. These types of words are often referred to as content words (the rest types of words, i.e., preposition, pronoun, determiner, conjunction, auxiliary, particle, are often referred to as grammatical words). However, it is worth noting that no grammatical requirements were set for the collocates. As a result, the final list may include both content-word collocates, and grammatical-word collocates. Examples of content-word collocates are *last **night***, *go **out***, and *very **nice***. Examples of grammatical-word collocates are *want **it***, *get **in***, and *so **much***. The ten million word spoken section of the BNC was used and the eligible pivot words were searched by the WordSmith Tool 3.0 and acquired the corresponding collocate(s). The final list contains 4,696 collocations. Table 4.1 shows the first 30 entries in Shin and Nation’s (2008) list.

<i>you know</i>	<i>I think (that)</i>	<i>a bit</i>	<i>used to {inf}</i>	<i>as well</i>	<i>a lot of</i>
<i>{No.} pounds</i>	<i>thank you</i>	<i>{No.} years</i>	<i>in fact</i>	<i>very much</i>	<i>{No.} pound</i>
<i>talking about</i>	<i>{No.} percent</i>	<i>I suppose</i>	<i>at the moment</i>	<i>a little bit</i>	<i>looking at</i>
<i>this morning</i>	<i>(not) any more</i>	<i>come on</i>	<i>number {No.}</i>	<i>come in</i>	<i>come back</i>
<i>have a look</i>	<i>in terms of</i>	<i>last year</i>	<i>so much</i>	<i>{No.} years ago</i>	<i>county council</i>

Table 4.1 The first 30 collocations in Shin and Nation's (2008) list

Durrant (2009) sought to generate a list of essential academic collocations for English for Academic Purposes (EAP) learners while defining “academic collocations” as “word pairs that co-occur with at least moderate frequency across a wide range of disciplines, but which are not often found in non-academic language” (p.161). In order to extract significant positionally-variable academic collocations that are pertinent across disciplines, Durrant first constructed an academic corpus that could “represent the probable diversity of collocation usage among disciplines” (2009:159). To achieve this, he gathered five million words from each of the 31 schools in the five university faculties where he teaches, including the School of English Studies. Finally, the corpus contained approximately 25 million tokens gathered from five disciplines, namely arts and humanities, life sciences, science and engineering, social-administrative and social-psychological. A list of word pairs meeting specific statistical criteria (namely, with a relative frequency greater than once per million and $MI \geq 4$ for word pairs within a four-word span of each other) was created for each cluster. The collocations appearing in each list were selected as common to all groups. To determine which collocations were exclusively academic, the keyness, operationalized as log-likelihood ratio, for each collocation common to all groups was computed from the original corpus and the BNC, representing academic and non-academic language, respectively. Then the collocations were ordered according to the difference of the log-likelihood ratio, yielding an ordered list of the most distinctively academic collocations, which were then manually vetted. Ultimately, the top 1000 items were included in the list. It was also

found that almost 75% of all the collocations (763 out of 1000) were grammatical, meaning they contained at least one non-lexical word. Table 4.2 lists the first 20 entries in Durrant's (2009) list.

<i>this study</i>	<i>associated with</i>	<i>this paper</i>	<i>based on</i>	<i>and respectively</i>
<i>due to</i>	<i>consistent with</i>	<i>between and</i>	<i>was performed</i>	<i>related to</i>
<i>compared to</i>	<i>was used</i>	<i>present study</i>	<i>number of</i>	<i>as shown</i>
<i>these results</i>	<i>respect to</i>	<i>to determine</i>	<i>note that</i>	<i>our study</i>

Table 4.2 The first 20 collocations in Durrant's (2009) list

Simpson-Vlach and Ellis (2010) extended previous research on lexical bundles by selecting 3-, 4-, and 5-word sequences that are valuable for language learners in academic contexts. They combined the Michigan Corpus of Academic Spoken English, BNC files of academic speech, and selected BNC academic writing files, totaling 2.1 million tokens, to identify lexical bundles that commonly occur in both written and spoken academic language. Using frequency and MI scores as measurements, they compiled three lists of valuable lexical bundles: the Academic Formulas List (AFL) of 207 frequently occurring academic lexical bundles across different academic disciplines, a list of the top 200 lexical bundles in spoken academic English, and a list of the top 200 lexical bundles in written academic English. Table 4.3 lists the first 20 entries in Simpson and Ellis's (2009) Academic Formulas List.

<i>[a/the] form of</i>	<i>(in) such a (way)</i>	<i>the distribution of</i>	<i>the problem of</i>	<i>(as) a function (of)</i>
<i>(in) terms of</i>	<i>the existence of</i>	<i>the process of</i>	<i>based on [a/the]</i>	<i>in which the</i>
<i>(the) extent to which</i>	<i>the question of</i>	<i>focus on the</i>	<i>is based on</i>	<i>(the) fact that</i>
<i>the role of</i>	<i>form of the</i>	<i>nature of the</i>	<i>the idea that</i>	<i>the structure of</i>

Table 4.3 The first 20 entries in Simpson and Ellis's (2009) Academic Formulas List

Furthermore, the researchers developed a "formula teaching worth" (FTW) metric. This is done by endowing FTW scores to 108 randomly selected academic lexical bundles that meet specific frequency criteria, with the FTW ranging from 1 to 5 to represent low to high teaching worth, respectively. They then used a correlation analysis, with both MI and frequency as coefficients, to determine which one better predicted the FTW. The correlation analysis results

showed that the MI score, which measures the degree to which the words are bound together in the lexical bundle, was the major determinant in measuring the pedagogical value of the sequences (teaching worth = β 0.56 MI + β 0.31 frequency).

Liu (2012) observed that the lexical bundles lists developed by Simpson-Vlach and Ellis contained some structurally incomplete Lexical bundles, such as *from of the* and *this is the*. To eliminate such incomplete lexical bundles, Liu created a more effective approach for identifying useful lexical bundles. This involved presenting the structurally-incomplete lexical bundles as partially-filled constructions, which included additional information such as word class. For example, *this is the* was presented as *this is det (determiner) + NP (noun phrase)*". Liu conducted a study to investigate the most common multiword constructions (MWCs), including lexical bundles, idioms, and phrasal/prepositional verbs, in a larger dataset consisting of the academic sections of the BNC and the Corpus of Contemporary American English (COCA). The selection criteria for MWCs included frequency (20 tokens per million in either corpus) and range distribution (20 tokens per million words in six out of the eight academic divisions in COCA or five out of six academic divisions in BNC). As the two corpora could only be searched through their supplied search engines and no download was available, Liu referred to existing MWCs found in the works of other researchers, such as Biber et al. (1999), Simpson-Vlach and Ellis (2010), and the Cambridge International Dictionary of Idioms (1998), and searched for each of them in the corpus interactive interface to determine whether they met the selection criteria. After manual adjustments, this resulted in a final list of the 228 most frequently used MWCs in academic written English. Table 4.4 lists the first 20 entries in Liu's (2012) list.

<i>such as (det + N)</i>	<i>for example</i>	<i>as well as (det + N)</i>	<i>NP suggests that</i>	<i>according to (det + N)</i>
<i>(be) based on</i>	<i>there be det + N</i>	<i>there be no NP</i>	<i>a/the number of</i>	<i>out of (det + N)</i>
<i>one of (det + N)</i>	<i>NP shows that</i>	<i>be/to be able to VP</i>	<i>focus on (det + N)</i>	<i>(as a) part of (det + N)</i>
<i>NP argue that</i>	<i>in addition (to)</i>	<i>(modal V) lead to</i>	<i>the fact that</i>	<i>(be) associated with</i>

Table 4.4 The first 20 entries in Liu's (2012) list

Prior research has sought to enhance the quality of multi-word collocation (MWC) lists by increasing their positional variability, applicability across multiple domains, and grammatical completeness. However, the value of such lists has remained limited due to their small size, typically comprising fewer than 1000 items. To address this issue, Ankermann and Chen (2013) developed a larger academic collocation list, which focused on two-word units, specifically lexical collocations where both words are lexical. They defined 'collocation' as a single word, or node word, that tends to co-occur within a span of ± 3 words from the node word. To compile a list of the most pedagogically relevant collocations, they utilized the Pearson International Corpus of Academic English, which contained 25 million words, and undertook a four-step process that involved computational analysis, refinement of the data-driven list based on quantitative and qualitative parameters, expert review, and systematization. Their computational analysis focused on collocations with the following grammatical relationships: verb + noun, adjective + noun, adverb + adjective, and adverb + verb. They began by generating a list of node words that appeared more than five times per million words and in at least five different texts, which they then used to extract potential collocations from the corpus. To extract collocations, they employed a combination of frequency and other statistical parameters, including a normed frequency of at least 1 per million, a normed frequency of at least 0.2 per million in each field of study, a mutual information (MI) score of at least 3, and a t-score of at least 4, following Simpson-Vlach and Alice (2010). The potential collocations filtered out in this phase were subject to an expert review process to ensure pedagogical relevance. These collocations were reviewed by six experts from the department of English linguistics in Coventry University to judge whether the potential collocations can be included in the final list. Ultimately, their final Academic Collocation List (ACL) comprised the 2,468 most pedagogically relevant lexical

collocations. Table 4.5, 4.6 and 4.7 lists the first 10 collocations of adj + noun, verb + noun, and adv + adj respectively in Ackermann and Chen's (2013) Academic Collocation List.

<i>academic writing</i>	<i>brief overview</i>	<i>causal link</i>	<i>conflicting interests</i>	<i>conventional wisdom</i>
<i>crucial factor</i>	<i>crucial importance</i>	<i>crucial role</i>	<i>cultural heritage</i>	<i>deep understanding</i>

Table 4.5 The top 10 adj + noun collocations in the Academic Collocation List

<i>achieve (a) goal</i>	<i>achieve (an) objective</i>	<i>cast doubt</i>	<i>make (a) living</i>	<i>make (a) prediction</i>
<i>meet criteria</i>	<i>make (a) requirement</i>	<i>obtain (a) result</i>	<i>pose (a) question</i>	<i>provide (a) clue</i>

Table 4.6 The top 10 verb + noun collocations in the Academic Collocation List

<i>ever increasing</i>	<i>hardly surprising</i>	<i>increasingly important</i>	<i>mutually exclusive</i>	<i>radically different</i>
<i>relatively few</i>	<i>relatively high</i>	<i>relatively little</i>	<i>relatively simple</i>	<i>slightly different</i>

Table 4.7 The top 10 adv + adj collocations in the Academic Collocation List

Hsu (2014) aimed to present a pedagogically useful list of the most frequent semantically non-transparent formulaic sequences (which were operationalized as 2-grams, 3-grams, 4 grams, and 5-grams) for non-English majors. The College Textbook Corpus (CTC), which is comprised of 200 college textbooks and totals 25 million tokens, was used and the n-gram function of *Collocate* was used to extract collocations. Hsu requested that the sequences included in the final list must occur with a minimal relative frequency of 5 times per million tokens, must occur in each of the 40 disciplines across 4 academic domains and at least 100 out of the 200 compulsory textbooks across 40 disciplines. The final list contains 475 formulaic sequences. It is worth noting that Hsu calculated the coverage of her opaque formulaic sequence list by dividing the number of tokens that the opaque formulaic sequence list occupies in the source corpus by the total number of tokens in the source corpus, resulting a coverage of 2.08%. Table 4.8, 4.9, 4.10 lists the first 10 entries of 2-word, 3-word, and 4-word opaque sequences respectively in Hsu's (2014) formulaic sequence list.

<i>appear to</i>	<i>a few</i>	<i>a little</i>	<i>a lot</i>	<i>at all</i>
<i>according to</i>	<i>account for</i>	<i>across from</i>	<i>in that</i>	<i>along with</i>

Table 4.8 The top 10 2-word opaque sequences in Hsu's (2014) formulaic sequence list

<i>(be) likely to</i>	<i>(be) supposed to</i>	<i>a number of</i>	<i>as well as</i>	<i>in addition to</i>
<i>in order to</i>	<i>in terms of</i>	<i>the rest of</i>	<i>with regard to</i>	<i>with regard to</i>

Table 4.9 The top 10 3-word opaque sequences in Hsu's (2014) formulaic sequence list

<i>as a result of</i>	<i>on the one hand</i>	<i>on the other hand</i>	<i>the extent to which</i>	<i>in the face of</i>
<i>in the course of</i>	<i>in the absence of</i>	<i>in the short turn</i>	<i>in the first place</i>	<i>not only... but also</i>

Table 4.10 The top 10 4-word opaque sequences in Hsu's (2014) formulaic sequence list

Rogers et al. (2021) conducted a recent academic study aimed at addressing the limitations of previous academic collocation lists. They critiqued Ackerman and Chen's ACL for not including frequency data of the collocations in the original corpus and for lacking example sentences to aid in contextual understanding, which they deemed necessary for language learners. Furthermore, they believed that the multiword units explored in previous research were insufficient in capturing the comprehensive integration of word groupings. Multiword units were defined as collocations operationalized as a *pivot word* (a high-frequency academic word that is used to search for other words which co-occur with it frequently) and a *collocate* (a word that commonly co-occurs with another word). To address these issues, they utilized the academic section of the COCA and Gardner and Davies' (2014) high-frequency academic vocabulary list to identify the most frequent 500 lemmas as pivot words for identifying lemmatized collocates. The concgramming method developed by Greaves (2009) was applied to search for collocates occurring within a span of four words to the left or right of each pivot word. They applied statistical parameters of one occurrence per million tokens and an MI score of at least three, followed by manual vetting to finalize the list. The resulting list consisted of 5057 academic multiword units, one of the largest to date, but it does not apply any grammatical requirements to the collocation selection process. Therefore, it may include some collocations that cannot be used in certain phraseological sophistication measures, like Paquot's (2019), that require specific

grammatical relationships between the collocates. Table 4.11 shows the first 20 entries in the collocation list by Rogers et al (2021).

<i>result suggest</i>	<i>variable relate</i>	<i>study examine</i>	<i>factor associate</i>	<i>indicate significant</i>
<i>often refer</i>	<i>may result</i>	<i>I argue</i>	<i>participate study</i>	<i>significant relationship</i>
<i>effect on</i>	<i>engage behavior</i>	<i>reflect difference</i>	<i>this article</i>	<i>determine extent</i>
<i>more likely</i>	<i>any attempt</i>	<i>likert range</i>	<i>on basis</i>	<i>high level</i>

Table 4.11 The top 20 entries in the collocation list by Rogers et al (2021)

4.2 Constructs and Operationalizations

Table 4.12 provides a comprehensive summary of the research objectives, corpora utilized, methodologies employed, statistical constructs applied, and the corresponding results obtained from previous studies.

Observations from Table 3.1 reveal that the construction of multiword lists often leads to the generation of larger lists, comprising several hundred (e.g., Shin & Nation, 2008) to several thousand entries (e.g., Ackermann & Chen, 2013; Roger et al, 2021). Moreover, recent research tends to conflate different statistical constructs in the process of developing multiword lists, developing from a sole emphasis on frequency (e.g., Shin & Nation, 2008) to a combination of frequency, association, and dispersion (e.g., Durrant, 2009; Ackermann & Chen, 2013).

Studies	The Corpora Used	Methodology	Constructs			Result
			Frequency	Association	Dispersion	
Shin & Nation (2008)	BNC spoken section with a size of approximately 10 million tokens.	Collocation is used to refer to a group of two or more adjacent words that occur frequently together. A collocation is made of two parts – a pivot word which is the focal word in the collocation and its collocate(s), the word or words accompanying the pivot word. The pivot words must occur in the most frequent 1,000 content words according to Leech, Rayson, and Wilson (2001) and were set as a noun, a verb, an adjective, or an adverb. The eligible pivot words were searched by the WordSmith Tool 3.0 and acquired the corresponding collocate(s).	Eligible collocations must occur minimally 3 times per million tokens (PMT) globally in the whole corpus.	No requirement.	No requirement.	The final list includes 4,698 collocations.
Durrant (2009)	An academic corpus which the author built exclusively for this research and is comprised of 25 million tokens of learner texts from various disciplines. Additionally, the non-academic sections of BNC	Academic collocations were defined as word pairs that co-occur at least moderate frequency across a wide range of disciplines, but which are not often found in non-academic language. They were operationalized as word pairs which co-occurred within a four-word span of each other.	Eligible collocations must have a minimal relative frequency (RF) of 1 time PMT globally.	Eligible collocations must have a minimal mutual information (MI) score of 4.	Eligible collocations must occur in each of the five sub-corpora.	The top 1,000 collocations with the largest difference of LL scores were included in the final list.

	was used as reference corpus.					
Simpson-Vlach & Ellis (2010)	Michigan Corpus of Academic Spoken English (1.7 million tokens) plus BNC files of academic speech (1.2 million tokens)	The Academic Formulas List includes formulaic sequences that are subsumed in 3-, 4-, and 5- phrases, identifiable as frequent recurrent patterns in written and spoken corpora that are significantly more common in academic discourse than in non-academic discourse and which occupy a range of academic genres. They were accordingly operationalized as 3-grams, 4-grams, and 5-grams.	Eligible formulaic sequences must have a minimal RF of 10 times PMT globally.	The authors combined the metrics of frequency and MI and created a "formula teaching worth". However, no specific MI score requirement was set.	No requirement.	The Academic Formulas List contains 207 core formulaic sequences with the highest FTW scores.
Martinez & Schmitt (2012)	BNC (100 million tokens)	The phrasal expressions were operationalized as 2-grams, 3-grams, and 4 grams in the BNC. The n-gram function of WordSmith Tools 5.0 was used to extract collocations.	The expressions in the final list must occur with a minimal RF of 0.05 times PMT globally.	No requirement.	No requirement.	The PRASE List includes 505 most frequent non-transparent multiword expressions.
Liu (2012)	Academic writing sections of the COCA and the BNC (totally 98.24 million tokens)	This study aimed to find the most common MWCs of a broad variety, including LBs, idioms, and phrasal/prepositional verbs. The operationalization for MWCs was done with the corpora search engine on https://www.english-corpora.org/ using the academic sections of BNC and COCA. Liu used	Eligible MWUs must have a minimal RF of 20 times PMT globally.	No requirement.	Eligible MWUs must appear in six out of the eight academic divisions in COCA or five out of six academic divisions in BNC.	The final list includes 228 most common MWCs.

		published sources as a database for query which includes lists of LBs proposed by previous researchers and lexicographers, e.g. Academic Formula List by Simpson-Vlach and Ellis (2010) and Oxford idioms dictionary (2001).				
Chon & Shin (2013)	The British Academic Spoken English Corpus (BASE) and the Academic Corpus (totally 5.1 million tokens)	Collocations were used to refer to a group of two or more adjacent words that occur frequently together. A collocation is made of two parts – a pivot word which is the focal word in the collocation and its collocate(s), the word or words accompanying the pivot word. In the research, the pivot words were set as the top 20 ranking academic words retrieved from each of the BASE and the Academic Corpus, and a pivot word must be either a noun, a verb, an adjective, or an adverb. The Concord function of the WordSmith Tools 3.0 was used to extract collocations.	The collocations included in the final list must occur with a minimal RF of 1.875 times PMT in the spoken corpus and a 1.741 times PMT in the written corpus.	No requirement.	No requirement.	The final list includes 934 written and 460 spoken collocations.
Ackermann & Chen (2013)	The written section of the Pearson International Corpus of Academic English	Collocation was defined as a single word (node word) that tends to co-occur in the span of ± 3 words from the reference word. MonoConc	The content pivot words must occur at least 5 times PMT and must	Eligible collocations must have a minimal MI score of 3 and a	Eligible collocations must appear in each field of study.	The ACL includes 2,468 collocations.

	(25.6 million tokens)	Pro 2.2 was used to obtain a list of content words in the corpus as node words. The list of node words was then used to extract collocations. This results in over 130,000 potential collocations, which were then lemmatized, and part-of-speech (POS) tagged. Finally, collocations with the following POS combinations were included in the final list: verb + noun, adjective + noun, adverb + adjective, and adverb + verb.	occur in at least 5 different texts. The collocations included in the final list must have a minimal RF of 1 PMT globally and a minimal RF of 0.2 in each in each field of study.	minimal t-score of 4.	Additionally, the RF in each field of study should be 0.2 or higher.	
Hsu (2014)	College Textbook Corpus (CTC) which is comprised of 200 college textbooks and totals 25 million tokens	The multiword sequences were operationalized as 2-grams, 3-grams, 4 grams, and 5-grams in the CTC. The n-gram function of <i>Collocate</i> was used to extract collocations.	The sequences included in the final list must occur with a minimal RF of 5 times PMT globally.	No requirement.	The sequences included in the final list must occur in each of the 40 disciplines and in at least 100 out of the 200 compulsory textbooks.	The final list includes 475 opaque formulaic sequences of 2-5 words.
Rogers et al (2021)	The academic section of the COCA, totally 83 million tokens	MWUs were operationalized as a pivot word and a collocate. The <i>AntWordPairs</i> were used to extract MWUs.	The MWUs included in the final list must occur with a minimal RF of 1 time PMT.	The MWUs included in the final list must have a minimal MI score of 3.	No requirement.	The final list includes 5,057 collocations that were judged as useful for academic learners.

Table 4.12 Summary of previous multiword unit lists

As pointed out clearly in the conclusion remarks Chapter 3, the construction of a new academic collocation list for research purposes, especially for examining L2 EFL writings is needed. It is crucial to determine the construct(s) incorporated into the new list. A closer examination of phraseological sophistication studies (e.g., Paquot, 2019; Vandeweerd et al, 2021; Rubin et al, 2021) revealed that phraseological sophistication was usually approximated by two methods. The first approach is to base on a list of sophisticated academic collocations as operationalized in Paquot (2019). The second approach is to base on the pointwise mutual information (PMI) scores of the three types of dependencies (i.e. adjectival modifiers, adverbial modifiers, and direct objects). The PMI score calculates the degree of association between the two elements in a collocation. In this way, the second method used to measure phraseological complexity is based on the construct of association. However, there is not an approach of approximating phraseological sophistication based on the constructs of frequency and dispersion. As a result, the compilation of the new academic collocation list will fill in this research gap by incorporating two constructs: frequency and dispersion.

Furthermore, few multiword unit lists were constructed on the lexis-grammar surface. The process of extracting these multiword units, as synthesized in Table 4.12, typically involves preliminarily establishing a list of node words. These node words were commonly set as content words as operationalized in Shin & Nation (2008) and Ackermann and Chen (2013). Subsequently, the extraction involves identifying collocates for these node words, often encompassing a span of four to five words on both the left and right side of node word as operationalized in Durrant (2009) and Ackermann & Chen (2013). However, few multiword unit lists were constructed on the lexis-grammar surface. Consequently, regulations governing the grammatical patterns of multiword units tend to be notably absent. However, as argued by many researchers, it is the collocations of certain grammatical patterns (e.g., collocations of verb +

noun) are particularly difficult for learners to master (e.g., Durrant & Schmitt, 2009; Bestgen & Granger, 2018). As a result, this new academic collocation list is designed on the surface of lexis-grammar and result in a list of academic collocations of certain grammatical relations.

In conclusion, this study aims to develop a new academic collocation list for the following reason.

1. The existing multiword unit lists have primarily been developed for pedagogical purposes, and there is a lack of dedicated English academic collocation lists designed specifically for researching L2 EFL learners' writings.
2. Few research has highlighted the challenges learners face in mastering collocations within certain grammatical relationships (e.g., verb + noun, adjective + noun) at the lexis-grammar interface, making these areas valuable for investigating L2 phraseological sophistication.
3. Furthermore, the size and coverage of existing academic collocation lists, such as the Academic Collocation List, is relatively limited.

In light of these gaps and challenges, this study aims to address them by constructing an academic collocation list, namely, the New Academic Collocation List, from a lexis-grammar perspective with a larger size and broader coverage for research purposes. After the compilation is completed, an analysis of learner texts will be conducted to examine the New Academic Collocation List. The study seeks to answer the following research questions:

- 1 Which collocations exhibit high frequency and consistent occurrence across academic disciplines?
- 2 How does the New Academic Collocation List compare to the Academic Collocation List (Ackermann and Chen, 2013) in terms of collocational coverage?

5. Data

This chapter focuses on the data utilized in the present research. In Section 4.1, the corpus employed for extracting academic collocations is introduced. Section 5.2 provides a detailed explanation of the pre-processing steps that were carried out on the corpus before conducting collocation extractions. This section also covers the statistical constructs used and their corresponding operationalizations. Additionally, in Section 5.3, the outcomes of the data extraction and calculation process are presented.

5.1 Data Description

At the onset of the study, a crucial initial step was to determine the academic corpus from which to extract academic collocations. In order to generate a comprehensive list of academic collocations, it was deemed necessary to utilize a corpus that was representative, current, and of substantial size.

Therefore, the author evaluated the use of existing academic corpora or academic sections from comprehensive corpora, such as the Michigan Corpus of Spoken Academic English (MICASE; Swales, Briggs, Ovens, & Simpson, 2007), Michigan Corpus of Upper-Level Student Papers (MICUSP; Arbor, 2009), Pearson International Corpus of Academic English (PICA; Ackermann, De Jong, Kilgarriff, & Tugwell, 2011), academic sections of the British National Corpus (BNC; Davies, 2004) and Corpus of Contemporary American English (COCA; Davies, 2008) for the purpose of extracting academic collocations. However, it was determined that these sources were not optimal for the extraction of formulaic sequences.

When delving into discussions surrounding corpora, a pivotal dimension of consideration revolves around their size. Two terms are frequently used: "token" and "type." Distinguishing these two terms helps obtaining a clearer understanding of the size of corpora. A token refers to

each individual occurrence of a word within a text, including repeated instances. It represents every instance of a word. On the other hand, a type refers to a unique word within a text. It represents the different words or forms found within a text, regardless of the number of times they occur. To illustrate this, let's consider the sentence "the cat catches the mouse". In this sentence, there are five tokens: *the*, *cat*, *catches*, *the*, *mouse*. However, there are only four types: *the*, *cat*, *catches*, and *mouse*. The token count reflects the total number of words in the sentence, while the type count represents the number of distinct words present.

Research on the extraction of formulaic sequences, such as studies by Ackermann & Chen (2013), Durrant (2009), Hsu (2014), Martinez & Schmitt (2012), and Simpson-Vlach & Ellis (2010), have shown that phraseology extraction is typically based on large academic corpora, which generally contain more than 10 million tokens. In contrast, the corpora such as MICASE, MICUSP, and the Academic Corpus, are quantitatively small, and therefore may not be suitable for the extraction of academic collocations. In addition, there are also numerous problems with other academic corpora, starting from the academic sections of comprehensive corpora, e.g., British National Corpus (BNC) and Corpus of Contemporary American English. Such corpora contained discourse that was collected years ago (BNC in 1994 and COCA in 2010) and they contain primarily English of a specific region, i.e., British English and American English respectively. Other existing academic corpora like PICAIE are accessible only through specific corpus platforms, such as Sketch Engine. These platforms limit the scope of available functions, and may not allow for critical processing operations necessary for this research, such as part-of-speech tagging and syntactic parsing.

The author conducted an internship from September 2022 to January 2023 with the objective of building a corpus for the extraction of potential collocations. Combining the results of the author's internship, the Louvain Corpus of Research Articles (LOCRA) was finally

selected as the database, which received extension during the internship of the author. The Louvain Corpus of Research Articles (LOCRA) is a corpus currently being developed at the Centre for English Corpus Linguistics (CECL) at UCLouvain that aims to provide a representation of expert academic discourse across various disciplines of social sciences and humanities. It comprises research articles from highly reputable and influential journals that were selected by experts in each respective field. The LOCRA includes research articles from peer-reviewed, top-rated journals in 11 disciplines, including anthropology, business, economics, education, law, literature, linguistics, medicine, political science, psychology, and sociology. Due to the wide range of disciplines in social science and humanities and texts' high quality ensured by expert selection, it was deemed as an ideal corpus as the database to extract potential collocations. Originally, the corpus contained approximately 1 million tokens per discipline. This size (approximately 1 million tokens per discipline and 11 million tokens in total) was considered relatively small in terms of collocation extraction compared to other larger corpora used in multiword unit extractions (e.g., the Pearson International Corpus of Academic English used to compile the Academic Collocation List contains approximately 25.6 million tokens). Therefore, the author decided to enlarge the LOCRA. During September 2022 and January 2023, the author conducted an internship, and part of the internship's objective was to enlarge the LOCRA so that it functions as an ideal basis to extract potential collocations. During the internship, the LOCRA received a "balanced" extension. This is done by collecting research articles from journals in the original version of LOCRA that were deemed as high-reputational and collecting research articles published from the year the original LOCRA was compiled (i.e., 2012) to the year the internship took place (i.e., 2022) and collecting the same number of research articles in each year from 2012 to 2022. After extension, the corpus comprises 18,147,623 tokens in total and approximately 2

million tokens per discipline, doubling the size of the original LOCRA. Table 4.1.1 shows the complete journals included in the LOCRA.

Anthropology	Education	Law	Literature	Linguistics
1. American Anthropology 2. Annual Review of Anthropology 3. Cultural Anthropology 4. Current Anthropology 5. Ethnology (2012) 6. Journal of the Royal Anthropological Institute	1. American Educational Research Journal 2. Instructional Science 3. International Journal of Educational Development 4. Journal of Educational Psychology 5. Journal of Teacher Education 6. Review of Educational Research 7. Studies in Educational Evaluation 8. Teaching and Teacher Education 9. European Journal of Psychology of Education	1. Antitrust Law Journal (2018) 2. International Journal of Constitutional Law 3. International Journal of Law, Policy and the Family	1. The Journal of Commonwealth Literature 2. Comparative Literature Studies 3. Chaucer Review 4. Postcolonial Text 5. English Literary Renaissance (2016) 6. Reformation and Renaissance Review	1. Applied Linguistics 2. Discourse Studies 3. Journal of Linguistics 4. Language in Society 5. Language Learning 6. Applied Language Learning 7. Applied Psycholinguistics 8. Bilingualism Language and Cognition 9. The Canadian Modern Language Review 10. ELT Journal 11. Foreign Language Annals 12. International Review of Applied Linguistics in Language Teaching 13. Journal of Second Language Writing 14. Language Awareness 15. Language Assessment Quarterly 16. Language Learning 17. Language Learning & Technology 18. Language Testing 19. Language Teaching Research 20. The Modern Language Journal 21. Second Language Research 22. Studies in Second Language Acquisition 23. System 24. TESOL Quarterly
Business	Psychology	Political Science and International Relations	Sociology	
1. Entrepreneurship, Theory and Practice 2. Journal of Business Venturing 3. Journal of Management Studies 4. The Leadership Quarterly	1. Behaviour Therapy 2. Biological Psychology 3. Cognition & Emotion 4. Developmental Psychology 5. Journal of Applied Psychology 6. Journal of Neuropsychology 7. Journal of Consulting and Clinical Psychology 8. Journal of Experimental Psychology - General 9. Journal of Occupational and Organizational Psychology 10. Journal of Personality and Social Psychology 11. Journal of Vocational Behavior 12. Psychological Review 13. Brain and Cognition	1. International Organization 2. Journal of Common Market Studies 3. Journal of Public Policy 4. Public Administration Review	1. American Journal of Sociology (2016) 2. The British Journal of Sociology 3. Community, Work & Family 4. Current Sociology 5. European Journal of Political Research 6. Party Politics 7. Population and Development Review 8. Population Studies	

Table 5.1 Journals included in LOCRA

To extract academic collocations, a portion of LOCRA comprising 9 disciplines of social sciences and humanities was selected. These disciplines include anthropology, business, education, law, linguistics, literature, psychology, political science and international relations, and social science. The disciplines of economics and medicine were excluded from this study. The rationale behind this decision stems from the type of learner corpora that are typically used to examine EFL learners' writing. Prominent examples of such corpora include the International Corpus of Learner English (Granger et al, 2020) and The Varieties of English for Specific Purposes dAtabase (VESPA; Paquot et al, 2022). These corpora primarily consist of reports and research papers within the domain of social sciences and humanities (i.e., business communication, linguistics, and literature). Consequently, academic collocations extracted from a corpus representing the social sciences and humanities are the most suitable for examining learner writing within this specific field.

Table 4.1.2 below shows the number of texts and the number of tokens in each discipline.

Anthropology (Token: 946,789) (Text: 120)	Education (Token: 887,761) (Text: 112)	Law (Token: 869,573) (Text: 102)
Literature (Token: 855,987) (Text: 117)	Linguistics (Token: 1,032,089) (Text: 114)	Business (Token: 858,757) (Text: 84)
Psychology (Token: 1,255,920) (Text: 154)	Political Science and International Relations (Token: 873,712) (Text: 109)	Sociology (Token: 925,083) (Text: 120)

Table 5.2 Number of tokens and texts in each discipline

The LOCRA contains a wide range of academic disciplines, each containing a substantial amount to texts, making it ideal for academic collocation extraction purpose for this research. The LOCRA is stored in the format of text files and each academic journal article is stored in one txt file.

This section reviews the data used to in this research. The Louvain Corpus of Research Articles (LOCRA) was finally selected as the database to extract potential collocations because of its relatively larger size and wide disciplines of social sciences and humanities. The next step involves pre-processing the LOCRA, serving as the first step of potential collocation extraction.

5.2 Data Pre-processing

With the corpus determined, the subsequent step involves pre-processing of the data. Notably, the previous chapter highlighted the limited research conducted on the lexis-grammar interface and its influence on collocations. However, it has been observed that collocations featuring specific grammatical relationships pose challenges for learners in terms of mastery (e.g., Granger and Bestgen, 2014; Bestgen and Granger, 2018). These studies have shed light on the difficulties learners encounter when attempting to acquire and utilize collocations with particular grammatical structures. Therefore, recognizing the significance of these grammatical relationships, this study acknowledges the importance of investigating collocations within the context of their associated grammatical patterns.

In recent research on assessing the English competency of EFL learners using phraseological units, there is a growing trend of incorporating part-of-speech (POS) tagging and parsing techniques prior to collocation extraction. An example of this approach is found in Paquot (2019), where the concept of phraseological complexity was introduced and examined using collocations extracted from French EFL learners. In this study, the collocations were not solely certain adjacent POS combinations, as in the work of Durrant and Schmitt (2009), but rather on POS combinations within specific syntactic relationships. Paquot (2019) specifically focused on three syntactic relationships based on Stanford typed dependencies (De Marneffe &

Manning, 2017). These relationships include adjectival modifiers, adverb modifiers, and direct objects. A detailed explanation of these relations can be found in Table 4.2.1.

Syntactic Relation	Description	Example
Adjectival modifier (<i>amod</i>)	An adjectival modifier of a noun phrase is any adjectival phrase that serves to modify the meaning of the noun phrase.	She eats red meat. (<i>amod</i> : red, meat)
Adverbial modifier (<i>advmod</i>)	An adverb modifier of a word is a (non-clausal) adverb or adverb-headed phrase that serves to modify the meaning of the word.	Genetically modified food. (<i>advmod</i> : modified, genetically)
Direct object (<i>dobj</i>)	The direct object of a VP is the noun phrase which is the (accusative) object of the verb.	She gave me a raise. (<i>dobj</i> : gave, rise)

Table 5.3 Description of *amod*, *advmod*, and *dobj*

In contrast to the traditional approach of extracting specific and adjacent part-of-speech (POS) combinations (e.g., Granger and Bestgen, 2014), the approach that combines syntactic relations and POS combinations offers improved accuracy in identifying the syntactic relation of extracted collocations. Traditional extraction based solely on POS combinations may include instances without a clear syntactic relation. Furthermore, this approach ensures that syntactic relations, which may not be captured by adjacent POS combinations, are not overlooked. To illustrate this, consider the following example:

- 1) This **has impact** on the result.
- 2) This **has great impact** on the result.

The distinction between *have impact* and *have great impact* in the provided examples showcases the significance of considering syntactic relations in collocation extraction. In the first sentence, *have impact* represents a verb + noun combination with a syntactic relation of direct object, making it suitable for selection as a quantified collocation. However, in the second sentence, the adjective *great* appears between *has* and *impact*, resulting in a verb + adjective +

noun combination (*have great impact*) rather than just *have impact*. If the extraction solely relies on POS combinations, the direct object syntactic relation (*have impact*) may be overlooked, leading to reduced accuracy.

To address this issue, the present study follows the approach suggested by Paquot (2018, 2019) and involves the extraction of three syntactic relations, namely adjectival modifiers (*amod*), adverbial modifiers (*advmod*), and direct object (*dobj*). By incorporating these syntactic relations alongside POS combinations, a more precise and comprehensive understanding of collocations can be achieved.

For preprocessing of the data, the selected 9 disciplines of the VESPA corpus underwent part-of-speech tagging and syntactic parsing. The processing of part-of-speech tagging and syntactic parsing for the entire LOCRA (Learner Oral Corpus of Restricted Academic Texts) was performed using the spaCy library in Python. spaCy is an open-source natural language processing library specifically designed for Python, offering a diverse range of tools and functionalities for text data processing and analysis. Subsequently, within the Python environment, the extraction of syntactic relations, namely *amod*, *advmod*, and *dobj*, was conducted.

Upon extracting the syntactic relations, the calculation of all the statistical metrics was carried out using an in-house Perl program. The preprocessing, extraction of collocations and the subsequent metric calculations were all facilitated by the assistance of Hubert Naets from the Centre de traitement automatique du langage of UCLouvain. The author expresses deep gratitude to Mr. Naets for his invaluable support throughout the process.

To ascertain the accuracy and reliability of the automatic part-of-speech tagging and syntactic parsing, a pilot study was undertaken. Specifically, a sample of 355 words from an academic journal article from LOCRA was subjected to manual part-of-speech tagging and

parsing. This approach aimed to evaluate the performance of the automated processes used, ensuring the reliability of the data for subsequent analysis and interpretation. The pilot study aimed to evaluate the reliability of using spaCy for POS tagging and parsing in a corpus linguistics analysis. A paragraph comprising 355 words and 266 dependencies was subjected to automated processing using spaCy. To validate the accuracy of the results, a manual verification was conducted, involving manual POS tagging and parsing of the paragraph. The manual inspection revealed 14 errors in the POS tagging process, indicating an accuracy rate of approximately 96%. Moreover, 12 mistakes were identified in the parsing, resulting in a parsing accuracy of around 95%. These findings demonstrate that spaCy provides a reliable tool for POS tagging and syntactic parsing in corpus linguistics analysis, with a high level of accuracy and overall performance. Table 5.4 illustrates the complete data preprocessing workflow.

	Tools
Lemmatization	spaCy
Part-of-speech tagging	
Parsing	
Extraction of dependency	Perl
Compute raw and relative frequencies for potential collocations (cf. section 5.3)	
Compute range for potential collocations (cf. section 5.3)	

Table 5.4 Preprocessing workflow

This section reviews the pre-processing procedures of the data. It results the extraction of the three syntactic dependencies, namely adjectival modifiers, adverbial modifiers, and direct object. The next step involves the construct(s) incorporated in this study and their detailed operationalizations.

5.3 Constructs for this Study

As discussed in Chapter 4, existing studies of phraseological sophistication are mostly based on association and operationalized as pointwise mutual information scores. There is a need

to conduct the studies of phraseological sophistication based on frequency and dispersion. The rest of this section introduces the operationalizations of these constructs in detail.

5.3.1 Frequency

Table 5.5 presents a comprehensive overview of the operationalizations employed in this research to quantify the frequency of potential collocations.

Operationalizations	Descriptions
Raw frequency in LOCRA	A collocation's total raw frequency in LOCRA across all disciplines.
Relative frequency in LOCRA	A collocation's relative frequency in LOCRA across all disciplines. It is presented as per million tokens and calculated by dividing the raw frequency of the collocation in the corpus by the total number of tokens in the corpus, multiplied by one million.

Table 5.5 Operationalizations of frequency

The construct of frequency is operationalized in two ways. The first is a potential collocation's raw frequency in the corpus. The second is a potential collocation's relative frequency across the whole corpus. The relative frequency is calculated by dividing the raw frequency of a potential collocation by the total number of tokens in the corpus, multiplied by one million. These two operationalizations provide a comprehensive measure of how many times a potential collocation appears in the LOCRA.

5.3.2 Dispersion

Table 4.2.4 presents a comprehensive overview of the operationalizations employed in this research to quantify the dispersion of potential collocations.

Operationalizations	Descriptions
Range	Number of disciplines in which a potential collocation appears

Table 5.6 Operationalizations of dispersion

The construct of frequency is operationalized in one way, namely the number of disciplines in which a potential collocation appears. The selected VESPA corpus contains 9 disciplines. Consequently, the dispersion construct range ranges from 1 to 9. The range operationalization of dispersion offers a measure of the distribution a potential collocation has among the 9 selected disciplines in the LOCRA.

6. Developing the New Academic Collocation List

This chapter offers a comprehensive explanation of the procedures involved in selecting appropriate statistical constructs and determining specific threshold values as parameters for the compilation of the New Academic Collocation List (NACL) in section 6.1. It outlines the step-by-step process followed to ensure the accuracy and relevance of the collocations included in the NACL. Additionally, section 6.2 provides a general review of the NACL, offering an overview of its content and significance. The chapter serves to provide readers with a clear understanding of the methodology employed and the resulting compilation of the NACL.

6.1 Filtering method

As illustrated in Chapter 5, the new academic collocation list designed specifically for research purposes will focus on the constructs of frequency and dispersion. The construct of frequency, operationalized as raw and relative frequency, ensures that the collocations included in the new list have relatively high frequency. The construct of dispersion, operationalized as range (i.e., the number of disciplines a potential collocation appears in), ensures that the potential collocation occurs in a wide range of disciplines selected from the LOCRA.

Determining a cut-off value for frequency is a crucial step in extracting collocations, as it helps establish a threshold for inclusion in the collocation inventory. Martinez and Schmitt (2012:302) highlight the need to set a limit for the list, even though frequency is a valid indicator of usefulness. However, studies differ in the frequency thresholds chosen, often without providing sufficient explanation for their selection. Nonetheless, some studies, such as Durrant (2009), Ackermann and Chen (2013), and Rogers et al. (2021), have adopted a relative frequency threshold of at least once per million tokens as a standard. In this research, the same relative

frequency threshold is utilized. This threshold ensures that the selected collocations have a sufficient frequency to be considered relevant and useful in academic writing.

To ensure that the collocations included in the NACL are representative across academic disciplines, additional criteria are applied. The potential collocations must occur in each of the nine disciplines covered in the LOCRA dataset. This is drawn from Durrant (2009) where he required the potential collocations included in his academic collocation list must occur in each of the five disciplines of the academic corpus where his collocation list is extracted from. This ensures that, according to Durrant (2009:162), “collocations are used across a range of academic disciplines”.

Furthermore, the frequencies of the filtered collocations within the LOCRA corpus are compared to a reference academic corpus, the Corpus of Academic Journal Articles (CAJA), which is a large balanced English corpus of academic journal articles with approximately 79 million tokens. Any collocation that meets the conditions of a frequency of over 1 per million tokens and occurrence in each discipline in LOCRA but has a raw frequency of less than 5 in CAJA is discarded for further analysis. By comparing the frequencies of the potential collocations in a large academic corpus, this ensures that the high-frequency nature of the potential collocations included in the final list is not by chance.

It is important to note that, given the focus on the construct of frequency in this study, no specific requirement is set for the association strength of the collocations. The emphasis is placed on identifying collocations with a high frequency of occurrence across academic disciplines.

In addition to the frequency parameters, ensuring that the collocations in the NACL are exclusively 'academic' is a critical aspect of this research. To achieve this, a comparison is made between the frequencies of the collocations in question and their frequencies in a comprehensive, non-academic corpus. In this study, the non-academic portion of the British National Corpus

(BNC) is used for this purpose. If a collocation's relative frequency in the non-academic portion of the BNC exceeds its relative frequency in LOCRA, it is deemed not exclusively academic and is subsequently discarded for further analysis. This criterion ensures that the collocations selected for inclusion in the NACL are primarily associated with academic contexts and are not overly prevalent in non-academic usage. By applying this comparison, the research aims to maintain the academic specificity of the collocations in the NACL. The complete workflow is illustrated in Table 6.1.

Procedure	Operationalization
Frequency	A potential collocation must have a relative frequency of at least once per million tokens.
Range	A potential collocation must occur in each of the 9 selected discipline of the LOCRA.
CAJA Reference corpus	A potential collocation must occur more than 5 times in the CAJA reference corpus
BNC Reference corpus	The relative frequency of a potential collocation must exceed its counterpart in the BNC.

Table 6.1 The potential collocations filtering workflow

Finally, the collocations were subjected to a qualitative review, wherein each entry was assessed independently to determine whether it met the inclusion criteria. The objective of this stage was to refine the list further by excluding the following types of entries, as per the guidelines provided by Ackermann and Chen (2013):

1. Linguistically incomplete units;
2. Combinations with concrete geographical and temporal references (e.g., European countries);
3. Combinations that are often hyphenated rather than written separately (e.g., non-organization);

At this stage, the filtering process has been completed, resulting in a refined selection of collocations that meet the established criteria. These collocations, which have successfully passed the frequency and academic specificity filters, are now included in the final version of the NACL.

6.2 Result

This section reports the result of applying the threshold for the frequency and dispersion constructs after extracting all the three types of potential collocations (i.e., *amod*, *advmod*, and *dobj*) in the LOCRA.

6.2.1 Composition

After undergoing a series of computational analyses and manual refinement, the New Academic Collocation List has been finalized, consisting of a total of 3,756 entries. The NACL includes three syntactic dependencies (i.e., adjectival modifiers, adverbial modifiers, and direct object), capturing various grammatical relationships within academic English. The distribution of entries across different syntactic categories, as well as examples illustrating each category, can be found in Table 6.2. For the complete list of potential collocations included in the NACL, please check the Appendix.

Dependencies	POS I	POS I	No of entries	Percentages	Examples
Adjectival modifiers	adj	n	1,497	39.9%	<i>high level, long term, present study</i>
Adverbial modifiers	adv	v	1,066	29.4%	<i>well understood, significantly differ</i>
	adv	adj	295	7.8%	<i>statistically significant, relatively high</i>
Direct object	verb	n	898	23.9%	<i>have effect, play role, provide evidence</i>
Total			3,756	100%	

Table 6.2 Overview of final academic collocations in POS combinations

A total of 3,756 collocations were selected and included in the New Academic Collocation List. These collocations are distributed across three syntactic dependency categories:

adjectival modifiers, adverbial modifiers, and direct objects. Among these categories, the adjectival modifiers group comprises the largest number of entries, with 1,500 collocations, accounting for 39.9% of the total. The adverbial modifiers group is the second largest, with 1,361 collocations, representing 29.4% of the total. The direct object group contains the fewest collocations, with 898 entries, making up 23.9% of the total. Although there is a slight variation in the distribution among the three groups, the distribution is relatively balanced, with no significant disparity in the number of collocations across the syntactic dependency categories.

6.2.2 Coverage

As discussed in the third chapter, an important objective of this study is to compile an academic collocation list that contains a high coverage in the source corpus. Therefore, it is essential to examine the coverage of the New Academic Collocation List in the corpus from which it is extracted (i.e., LOCRA).

In the early days when researchers compiled single-vocabulary lists (i.e., word lists whose entries are all single words, e.g., Academic Word List; Coxhead, 2000), the dimension of coverage was operationalized by computing the portion of the occurrences all on-list vocabularies in the source corpus on the total token of the source corpus. For instance, Coxhead achieved an approximate 10% coverage of the Academic Word List by computing all the occurrences of the 570 word families that are on the Academic Word List in the source corpus (i.e., The Academic Corpus) by the total token of the source corpus. However, the method used to compute coverage for single-vocabulary lists may not be suitable for multiword unit lists, as highlighted in previous studies such as Ackermann and Chen (2013) and Hsu (2014). In the case of word combination lists, a different approach is needed to accurately assess their coverage.

For example, when calculating the coverage of a list of collocations of adjectival modifier syntactic dependency, it is crucial to consider the occurrences of the specific collocations on the

list in relation to all occurrences of multiword units with the same syntactic dependency in the entire source corpus. This ensures that the computation method aligns with the nature of multiword lists, which focus on specific word combinations rather than individual words. By adopting this alternative approach, the coverage analysis of multiword lists can provide a more accurate representation of the presence and utilization of these word combinations in the corpus under investigation. It takes into account the specific syntactic or collocational patterns and provides insights into how frequently the listed combinations occur within the larger context of similar combinations in the corpus. The formula below shows how the coverage of the New Academic Collocation List in the LOCRA is computed.

$$\frac{\text{The occurrences of the NACL dependencies in the LOCRA}}{\text{The occurrences of all the } amod, advmod, \text{ and } dobj \text{ dependencies in the LOCRA}} \times 100\%$$

The occurrences of the NACL collocations in the LOCRA are shown in Table 6.3:

	<i>amod</i>	<i>advmod</i>	<i>dobj</i>	Total
Occurrences	182,357	120,134	99,375	401,866

Table 6.3 Occurrences of the NACL in the LOCRA

The occurrences of all three types of potential collocations of different syntactic relations (i.e., *amod*, *advmod*, and *dobj*) in the LOCRA are shown in Table 6.4:

	<i>amod</i>	<i>advmod</i>	<i>dobj</i>	Total
Occurrences	1,444,219	490,926	705,534	2,640,679

Table 6.4 Occurrences of all three types of potential collocations in the LOCRA

The coverage of the New Academic Collocation List, computed using the proposed method, amounts to approximately 15.21% ($401,866 / 2,640,679$) of the LOCRA. To the best of the author's knowledge, this approach for examining the coverage of word combination lists has

not been previously utilized in any other research. Table 6.5 shows the individual coverage of the NACL *amod*, *advmod*, and *dobj* collocations in the LOCRA.

	<i>amod</i>	<i>advmod</i>	<i>dobj</i>
Coverage	$\frac{182,357}{1,444,219} = 12.6\%$	$\frac{120,134}{490,926} = 24.5\%$	$\frac{401,866}{2,640,679} = 15.2\%$

Table 6.5 individual coverage of the NACL *amod*, *advmod*, and *dobj* collocations in the LOCRA.

In contrast, when employing the traditional approach, which involves dividing the occurrences of all collocations on the list by the total number of tokens in the source corpus (without considering occurrences of collocations with the same syntactic dependency), the coverage of the New Academic Collocation List is approximately 2.37%.

7. Discussion

RQ1: Which collocations occur frequently and uniformly across academic disciplines?

The New Academic Collocation List (NACL), focusing on the construct of frequency and dispersion, comprises a total of 3,756 collocations. Within the NACL, there are 1,497 collocations with adjectival modifier syntactic dependency, 1,361 collocations with adverbial modifier syntactic dependency, and 898 collocations with direct object syntactic dependency. A closer examination of each group sheds light on the specific characteristics of the NACL and its composition.

To gain a deeper understanding of the NACL, the Academic Keyword List (ALK; Paquot, 2010) was utilized. This list consists of 930 words that are frequently used in a variety of academic texts, while occurring relatively infrequently in other types of text. The elements of the collocations in the NACL will be analyzed whether they are included in the AKL. By incorporating the Academic Keyword List into the analysis, the NACL can capture collocations that align with the vocabulary and linguistic patterns commonly found in academic discourse.

Collocations of the *amod* syntactic dependency

The top 30 collocations of the adjectival modifier syntactic dependency along with the highest relative frequencies are listed in Table 7.1.

Adj	Noun	Relative Frequency per million tokens	Adj	Noun	Relative Frequency per million tokens
high	level	189.5	high	school	49.6
long	term	113.3	middle	class	46.0
other	word	109.2	different	type	45.7
other	hand	107.2	political	party	44.6
same	time	107.1	other	country	40.8
present	study	103.0	large	number	40.7
human	right	92.56	social	capital	39.5
low	level	90.1	important	role	39.4
significant	effect	81.6	negative	effect	38.9
significant	difference	70.0	different	level	36.2

previous	study	65.5	wide	range	36.0
previous	research	60.0	public	policy	35.0
current	study	59.8	other	study	34.7
short	term	49.7	other	factor	34.0

Table 7.1 Top 30 *amod* collocations with the highest relative frequencies

Upon comparing the top 30 *amod* collocations with the highest frequency in Table 7.1 to previous multiword unit lists, it becomes evident an overlap exists between the *amod* collocations in the NACL with the Academic Formulas List (Simpson-Vlach and Ellis, 2010). Specifically, certain *amod* collocations listed in Table 7.1 are constituents of lexical bundles in the Academic Formulas List. For example, the *amod* collocations *other word*, *other hand*, *same time*, and *large number* are components of the lexical bundles *in other words*, *on the other hand*, *at the same time*, and *a large number of* respectively in the Academic Formulas List.

An additional analysis was conducted to examine the academic nature of the adjectives and nouns of *amod* collocations in the New Academic Collocation List. The findings revealed that the adjectives *other* and *different*, both of which were listed as academic keywords in the Academic Keyword List, formed the highest number of *amod* collocations, with 79 and 77 collocations respectively. Together, these two adjectives accounted for approximately 10% of all *amod* collocations. Following *other* and *different*, the adjectives *social* and *new*, which are also included in the Academic Keyword List, ranked next in terms of the number of *amod*

collocations. Regarding the nouns, the noun *way* emerged as the noun with the highest number of *amod* collocations, totaling 30 and representing approximately 2% of all *amod* collocations.

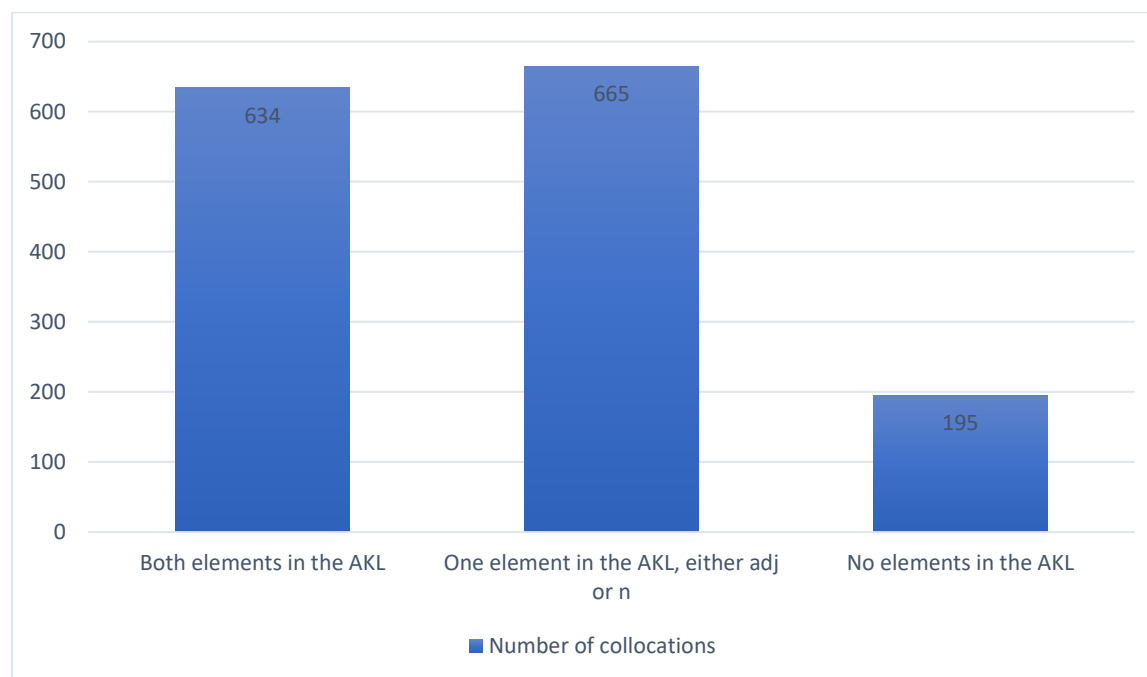


Figure 7.1 *amod* collocations' components in the AKL

Figure 7.1 illustrates the distribution of *amod* collocations based on their elements' inclusion in the Academic Keyword List. The figure presents three categories: collocations where both the adjective and the noun are included in the AKL, collocations where only one element (either the adjective or the noun) is included in the AKL, and collocations where neither the adjective nor the noun is included in the AKL. Figure 7.1 clearly demonstrates that the majority of *amod* collocations (87%) consist of at least one element that is included in the AKL. Only a small proportion, approximately 13% of *amod* collocations, are composed of words that are not considered key words according to the AKL.

Examples of collocations of the *amod* dependency whose both elements are included in the AKL are *high level*, *significant effect*, *previous research*, *wide range*, *total number*, *high quality*, *clear evidence*. Examples of collocations of the *amod* dependency whose only one

element (whether the adjective or noun) is included in the AKL are *future study*, *high cost*, *new framework*, *certain area*, *small group*, *such case*, *direct relationship*, *same kind*. Examples of collocations of the *amod* dependency that does not contain AKL vocabularies are *young people*, *private sector*, *many years*, *public school*. The words marked bold shows that they appear in the AKL.

It is worth noting that the *amod* collocations that does not contain AKL vocabularies contain a relatively higher level of association. The average MI score for all *amod* collocations is 1.94 whereas *amod* collocations that does not contain AKL vocabularies have an average MI score of 2.89. Some typical collocations with high MI scores are *informed consent*, *horizontal axis*, *very beginning*, and *violent crime*. The elements of these collocations often appear together as collocations. These collocations have relatively high frequency, wide range distribution in academic disciplines, and high association, which happen to align with the nature of the Academic Collocation List (Ackermann and Chen, 2013). The author argues that despite that neither element of these collocations appear in the AKL, they contribute to the academic discourse for their high association and wide range distribution in academic disciplines.

Collocations of the *dobj* syntactic dependency

The top 30 collocations of the direct object syntactic dependency along with the highest relative frequencies are listed in Table 7.2.

Verb	Noun	Relative frequency per million tokens	Verb	Noun	Relative frequency per million tokens
have	effect	144.7	solve	problem	30.2
play	role	131.8	make	use	28.6
take	place	105.0	draw	attention	27.6
have	impact	64.9	raise	question	27.5
provide	evidence	61.4	find	evidence	27.0
have	implication	50.1	shed	light	26.9
make	sense	45.5	ask	question	26.5
make	decision	43.5	show	effect	26.4

provide	support	40.4	address	question	26.1
test	hypothesis	40.0	use	model	25.9
provide	information	39.2	have	level	25.7
pay	attention	38.6	use	measure	25.6
answer	question	38.5	give	rise	25.5
examine	effect	33.1	have	potential	25.2
provide	insight	32.3	have	access	25.1

Table 7.2 Top 30 *dobj* collocations with the highest relative frequencies

It has been observed that some collocations of the top 30 *dobj* collocations with the highest relative frequencies requires the use of an adjacent preposition. For instance, *pay attention (to)*, *have impact (on)*, *shed light (on)*, and *give rise (to)*.

The New Academic Collocation List includes a substantial number of *dobj* collocations with the verb *have* as the verb. These collocations, totaling 108, account for approximately 12% of all *dobj* collocations. Additionally, the nouns *relationship* and *way* contribute significantly to the formation of *dobj* collocations, with a combined total of 40 collocations, representing approximately 4.5% of all *dobj* collocations. Examples of *dobj* collocations with *have* as the verb include *have effect*, *have impact*, *have access*, *have interest*, and *have association*. It is interesting to note that the verb *have* is not included in the Academic Keyword List.

This finding is intriguing in light of previous research on verb + noun collocations in learner corpus studies (e.g., Granger and Bestgen, 2014), which has shown that EFL learners face challenges in acquiring collocations with high association strength and low frequency. Collocations such as *provide information*, which involve verbs that are less common in comparison to more frequently used verbs like *give* and *take*, are particularly difficult for learners. However, the presence of a substantial number of collocations with the verb *have* which is relatively common highlights their value for EFL learners. It demonstrates that collocations

with common verbs, despite their higher frequency and wider usage, are still relevant and beneficial for learners to acquire.

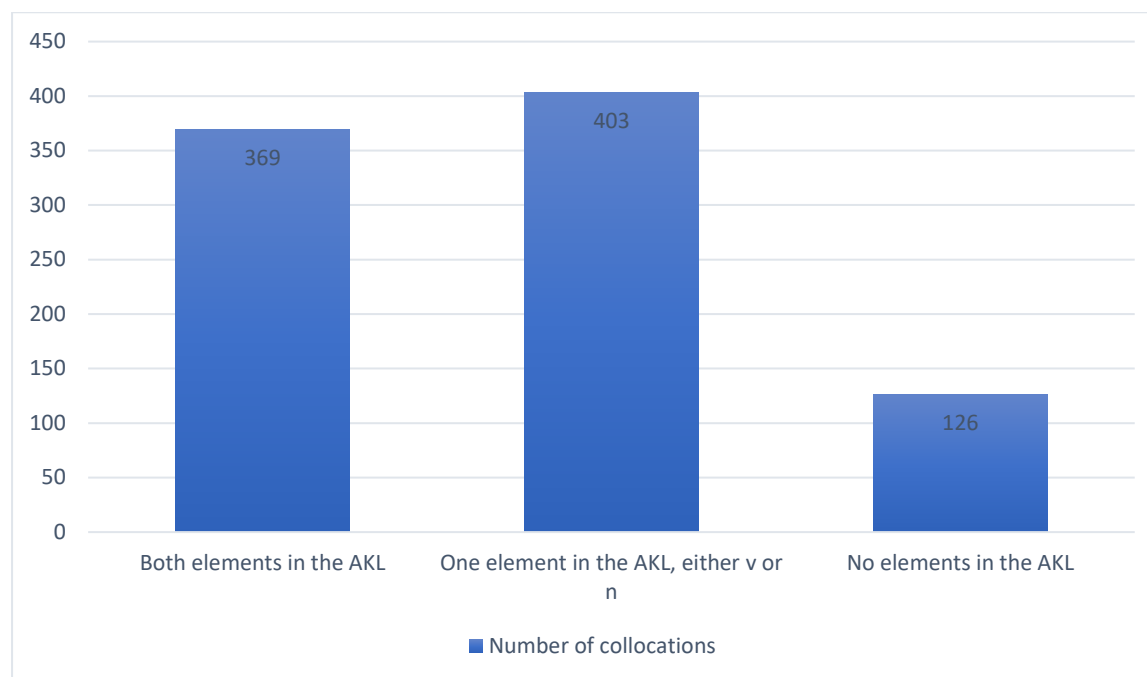


Figure 7.2 *dobj* collocations' components in the AKL

Figure 7.2 illustrates the distribution of *dobj* collocations based on their elements' inclusion in the AKL. The three categories include collocations with both the verb and the noun listed in the AKL, collocations with either the verb or the noun listed, and collocations with neither the verb nor the noun listed in the AKL. The analysis reveals that the majority of *dobj* collocations, approximately 86%, contain at least one element (either the verb or the noun) that is included in the AKL. This indicates that the AKL captures a significant portion of the key lexical items associated with *dobj* collocations. Conversely, only around 14% of *dobj* collocations do not involve words that are considered key words by the AKL.

Examples of collocations of the *dobj* dependency whose both elements are included in the AKL are *show relationship*, *support argument*, *gain understanding*, *consider impact*. Examples of collocations of the *dobj* dependency whose only one element (whether the verb or noun) is

included in the AKL are **conduct** *study*, **provide** *education*, **follow** *lead*, **reflect** *way*, *draw attention*, *give rise*, *have problem*, *deserve attention*. Examples of collocations of the *doj* dependency that does not contain AKL vocabularies are *find way*, *make effort*, *make statement*, *invest time*. The words marked bold shows that they appear in the AKL.

A closer examination of these collocations in the LOCRA suggests that they are a part of a common syntactic construction:

Verb + noun + preposition + verb + noun

In this construction, the bold verb + noun represents the collocation. Its usage is usually followed by another proposition, which is then followed by another verb. An example is shown below.

— *The intention **invest time to support** the implementation of change is...* (PSY019-11)

In this construction, there are two *doj* collocations, namely *invest time*, and *support implementation*. The first collocation functions as a transition to connect another *doj* collocation. In the NACL, many collocations of the *doj* dependency that does not contain AKL vocabularies function in this way. Although neither element of the collocation appears in the AKL, they contribute to the academic discourse by functioning as a transition to introduce another *doj* potential collocation.

Collocations of the verb + adverb syntactic dependency

The top 30 collocations of the verb + adverb syntactic dependencies with the highest relative frequencies are listed in Table 7.3.

Verb	Adv	Relative frequency per million tokens	Verb	Adv	Relative frequency per million tokens
do	so	134.3	differ	significantly	21.1
see	also	95.9	have	thus	20.4
be	however	42.4	suggest	however	19.3
find	also	42.3	use	often	19.2

know	well	42.2	establish	well	18.9
include	also	38.0	perform	well	18.8
call	so	36.6	make	also	18.5
show	also	36.3	discuss	above	18.4
understand	well	30.7	mention	above	17.7
use	also	29.4	fit	well	17.3
suggest	also	28.6	show	however	17.2
have	however	26.7	describe	above	16.9
provide	also	26.0	go	far	16.8
take	together	25.4	consider	also	16.7
find	however	23.3	present	here	16.0

Table 7.3 Top 30 *verb + adverb* collocations with the highest relative frequencies

An analysis of the top 30 verb + adverb collocations with the highest relative frequencies reveals distinct patterns. Adverbs such as *also*, *however*, and *well* appear frequently, forming 9, 5, and 5 collocations, respectively, within the top 30 collocations of verb + adverb syntactic dependency. Upon comparison with the Academic Formulas List, it becomes apparent that these specific collocations are unlikely to be included as component parts of the academic formulas listed in the Academic Formulas List. This disparity can be attributed to the different methods employed in compiling the two lists. The Academic Formulas List utilized n-gram extraction, which primarily captures strings of adjacent words. In contrast, the NACL is compiled based on parsed data, allowing for more precise identification of verb + adverb collocations. This distinction allows the NACL to offer new insights into the most frequent academic collocations within the verb + adverb syntactic dependency.

The comprehensive coverage of verb + adverb collocations in the NACL, beyond strictly adjacent word combinations, adds a valuable dimension to the study of academic language usage. It provides a more nuanced understanding of the specific verb + adverb collocations that are prevalent within academic discourse, highlighting their distinctiveness and significance.

Among collocations with the verb + adverb syntactic dependency, the verb *have* stands out as the most frequent, forming a total of 30 collocations. This observation aligns with the findings in *dobj* syntactic dependency, where the same verb *have* also generates the highest number of collocations. Following *have*, the verbs *use*, *consider*, and *take* emerge as the second, third, and fourth most productive verbs, forming 29, 22, and 22 collocations respectively.

In terms of adverbs, *also* dominates the list, generating the highest number of collocations with a total of 164, accounting for 15.4% of all verb + adverb collocations. The adverbs *however*, *thus*, and *often* follow closely behind, forming the second, third, and fourth highest number of collocations with totals of 86, 81, and 48 respectively.

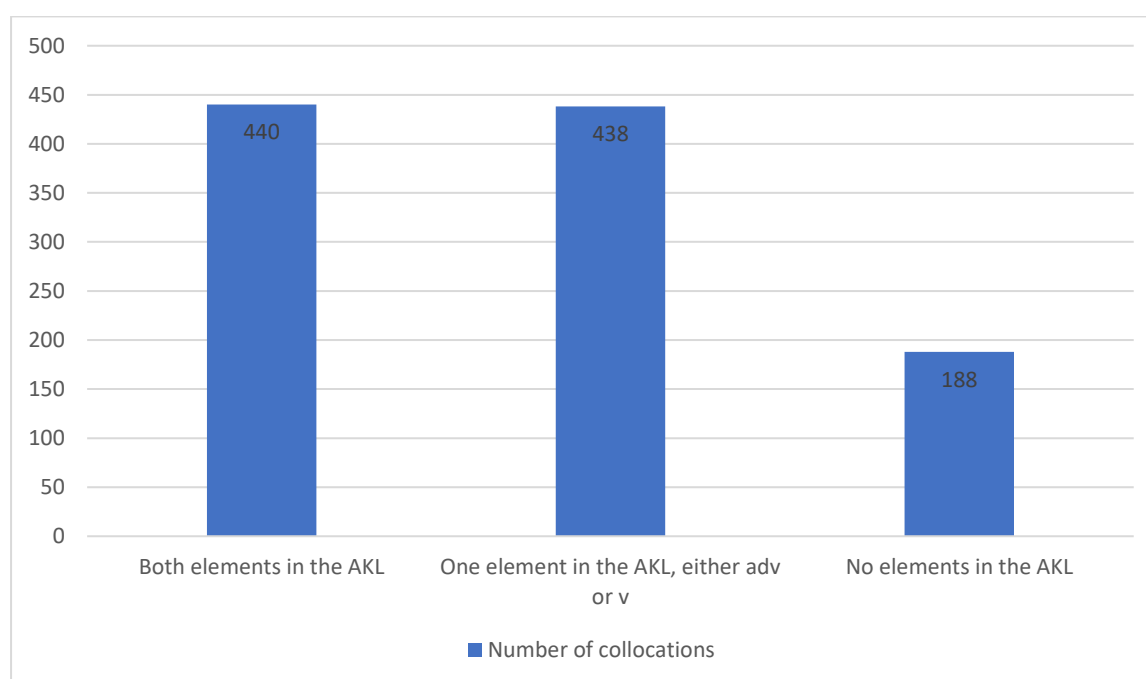


Figure 7.3 *verb + adverb* collocations' components in the AKL

Figure 7.3 illustrates the distribution of verb + adverb collocations across three categories: collocations where both the adverb and the verb are included in the AKL, collocations where only one element (either the adverb or the verb) is included in the AKL, and collocations where neither the adverb nor the verb are considered key words in the AKL and are therefore included in the

ACL. The figure highlights that the majority of verb + adverb collocations (82%) contain at least one element that is included in the AKL, indicating their recognition as key words in academic language. Only a minority (18%) of verb + adverb collocations are formed by words that are not classified as key words in the AKL. This demonstrates the strong presence and influence of the AKL in capturing the majority of verb + adverb collocations in the academic domain.

Examples of collocations of the adverb + verb dependency whose both elements are included in the AKL are ***also use**, **significantly differ**, **thus become**, **therefore include***. Examples of collocations of the adverb + verb dependency whose only one element (whether the adverb or verb) is included in the AKL are ***first** provide, **immediately** follow, **finally** show, **broadly** define, **however** **make**, **far** **explore**, **also** **offer**, **thus** **find***. Examples of collocations of the adverb + verb dependency that does not contain AKL vocabularies are *long stand, seriously take, negatively impact, still hold*. The words marked bold shows that they appear in the AKL. However, no specific pattern was observed for collocations of the *adverb + verb* dependency that does not contain AKL vocabularies.

Collocations of the adverb + adjective syntactic dependency

The top 30 collocations of the adverb + adjective syntactic dependencies with the highest relative frequencies are listed in Table 7.4.

Adv	Adj	Relative frequency per million tokens	Adv	Adj	Relative frequency per million tokens
more	likely	212.5	relatively	high	17.0
statistically	significant	82.4	more	general	16.8
less	likely	60.7	most	common	16.5
most	important	49.1	too	much	16.4
more	important	35.1	open	ended	15.6
more	complex	33.2	more	accurate	15.5
more	difficult	30.1	only	few	15.1
very	different	27.0	particularly	important	14.9
more	effective	26.5	very	similar	14.2
more	positive	23.9	relatively	small	14.1

significantly	high	23.4	very	few	14.0
most	likely	23.3	more	negative	13.5
more	recent	19.8	very	high	13.3
significantly	different	19.7	more	detailed	12.9
most	recent	17.2	more	common	12.4

Table 7.4 Top 30 *adverb + adjective* collocations with the highest relative frequencies

Table 7.4 highlights the patterns observed within the top 30 *adverb + adjective* collocations. It is evident that the distribution of adverbs is primarily centered around *more*, *most*, *less*, and *very*, with 21 out of the 30 collocations being formulated by these adverbs, which are commonly referred to as adverbs of degree. These adverbs serve to intensify or modify the degree conveyed by the adjectives. Among all *adverb + adjective* collocations, the adverb *more* appears most frequently, with a total of 133 collocations, accounting for 45.1% of the collocations. The adverbs *most*, *very*, and *relatively* rank second, third, and fourth respectively, with 31, 15, and 14 collocations. In terms of adjectives, *different* forms the highest number of collocations (14), followed by *high*, *important*, and *largest*, with 9, 9, and 7 collocations respectively.

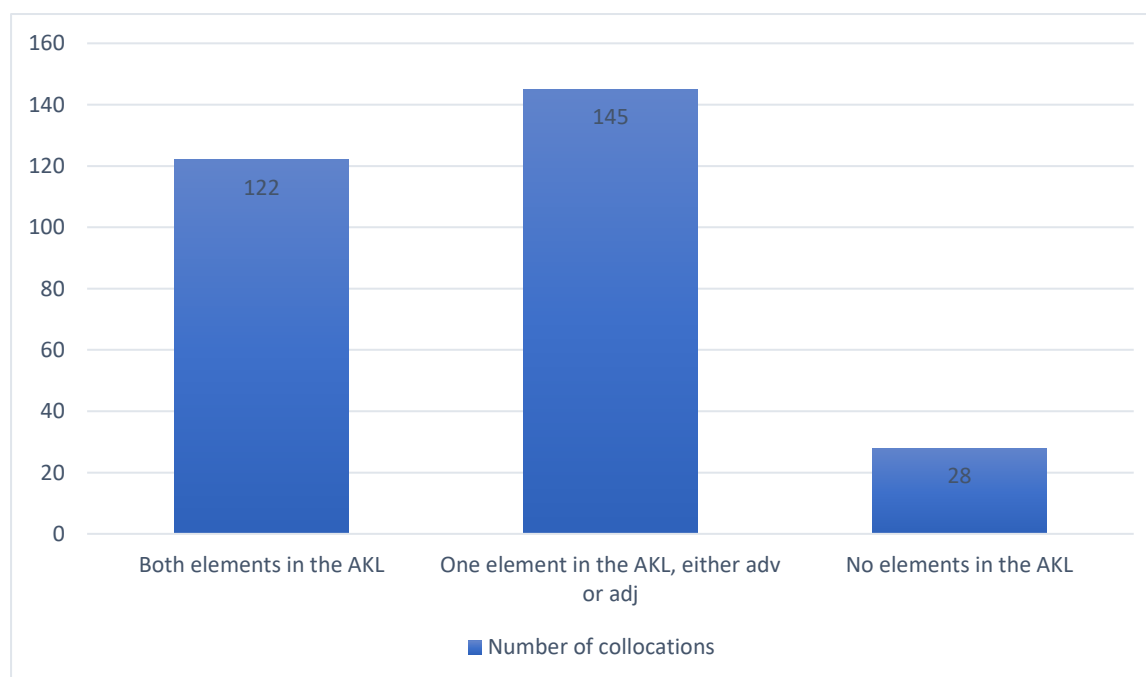


Figure 7.4 *adverb + adjective* collocations' components in the AKL

Figure 7.4 illustrates the distribution of three types of adverb + adjective collocations: those in which both the adverb and the adjective are included in the AKL, those with only one element included in the AKL (either the adverb or the adjective), and those in which neither the adverb nor the adjective are included in the AKL. The figure highlights that the majority of adverb + adjective collocations contain at least one element that is considered a key word according to the AKL (90%), while only a small portion (10%) consists of words that are not considered key words by the AKL.

Examples of collocations of the adverb + adjective dependency whose both elements are included in the AKL are ***significantly high**, relatively high, particularly relevant, readily available*. Examples of collocations of the adverb + adjective dependency whose only one element (whether the adverb or adjective) is included in the AKL are *unusually high, radically different, **entirely** new, hardly surprising, significantly **more**, relatively few, more saliant, particularly **strong***. Examples of collocations of the adverb + adjective dependency that does not contain AKL vocabularies are *mutually exclusive, slightly more, comparatively little*. The words marked bold shows that they appear in the AKL. However, no specific pattern was observed for collocations of the adverb + adjective dependency that does not contain AKL vocabularies.

RQ2: How does the New Academic Collocation List compare to the Academic Collocation List?

At this stage, it is critical to compare the New Academic Collocation List (NACL) with the Academic Collocation List (ACL, Ackermann and Chen, 2013) to see the similarities and differences of these two lists to gain insights into the improvements made in the new list for examining L2 phraseological sophistication.

The first improvement is that the NACL is a larger list of collocations in terms of total number of entries. The NACL contains 3,756 entries whereas the ACL contains 2,496 entries, a

1261 entries difference, which is approximately half number of the total collocations included in the ACL. The NACL, the larger list compared to the ACL means a more comprehensive inventory of phraseological resources representative of academic English. By including a greater number of entries, the NACL provides researchers with a more extensive pool of collocations to explore and utilize in various academic search.

In addition to the larger size, the NACL also has a larger coverage from various aspects. The traditional method of computing coverage by dividing the occurrences of on-list word combinations by the total tokens of the corpus is problematic for word combination lists. Instead, the NACL adopts a more appropriate approach by comparing the occurrences of on-list word combinations to the corresponding categories of word combinations in the corpus. In the case of the three syntactic dependencies (adjective modifiers, adverbial modifiers, and direct objects) in the LOCRA corpus, the total occurrences amount to 2,640,679. The NACL, on the other hand, has a total occurrence of 401,866, which accounts for approximately 15.2% of all collocations of these syntactic dependencies in LOCRA. In contrast, the ACL has a total occurrence of 157,716 in the LOCRA, representing approximately 6.0% of all collocations in the same syntactic dependencies in the LOCRA. The calculation of the coverage of the NACL and the ACL in the LOCRA is computed in the following formulas.

$$\frac{\text{Occurrences of the NACL collocations in the LOCRA}}{\text{Total occurrences of all the } amod, advmod, \text{ and } dobj \text{ dependencies in LOCRA}} = \frac{401,866}{2,640,679} = 15.2\%$$

$$\frac{\text{Occurrences of the ACL collocations in the LOCRA}}{\text{Total occurrences of all the } amod, advmod, \text{ and } dobj \text{ dependencies in LOCRA}} = \frac{157,716}{2,640,679} = 6.0\%$$

This comparison reveals that the NACL has approximately 3.5 times the coverage of the ACL in the LOCRA. Furthermore, if the coverage of the NACL is calculated using the method employed by Ackermann and Chen for the ACL, i.e., dividing the occurrences of all on-list collocations by the total number of tokens in the source corpus, the NACL demonstrates a

coverage of approximately 2.38%. In comparison, the ACL has a coverage of 1.4% using the same method. This indicates that the NACL approximately doubles the coverage of the ACL when measured in this way.

The third improvement of the New Academic Collocation List (NACL) is its relatively balanced distribution of collocations across the three syntactic dependencies. Among the 3,756 collocations in the NACL, 1,498 collocations belong to the syntactic dependency of adjectival modifiers, 1,361 collocations belong to adverbial modifiers, and 898 collocations belong to direct objects. This balanced distribution ensures that each syntactic dependency is adequately represented in the NACL. In contrast, the Academic Collocation List (ACL) exhibits a significantly different distribution. Among the 2,496 collocations in the ACL, the largest proportion comprises adjective + noun collocations, totaling 1,773 collocations and accounting for 71% of all entries. This is followed by adverb + adjective collocations and adverb + verb collocations, which together account for 28% of all entries in the ACL. Collocations of noun + noun and verb + adjective represent the smallest categories, with 62 and 30 collocations respectively. It is evident that the ACL is heavily skewed towards adjective + noun collocations, making up approximately three quarters of the list.

In comparison, the NACL demonstrates a more balanced distribution across the three syntactic dependencies. This inclusion of a more equitable representation of adjectival modifiers, adverbial modifiers, and direct objects allows for a more comprehensive portrayal of phraseological patterns in academic English. The distributions of collocations across various

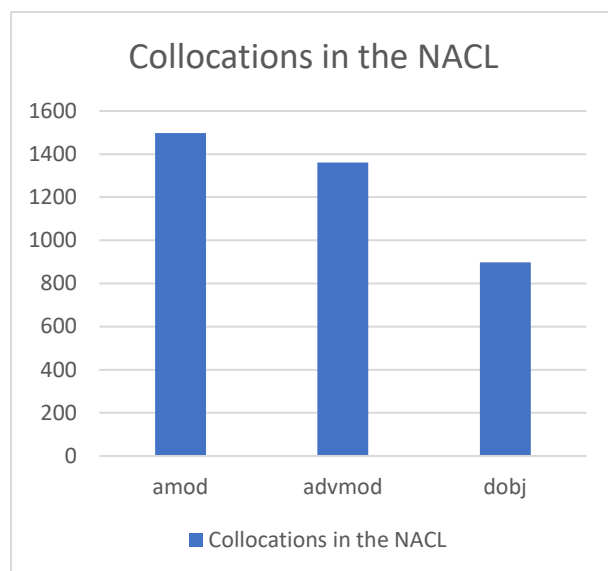


Figure 7.5 Distribution of categories of collocations in the NACL

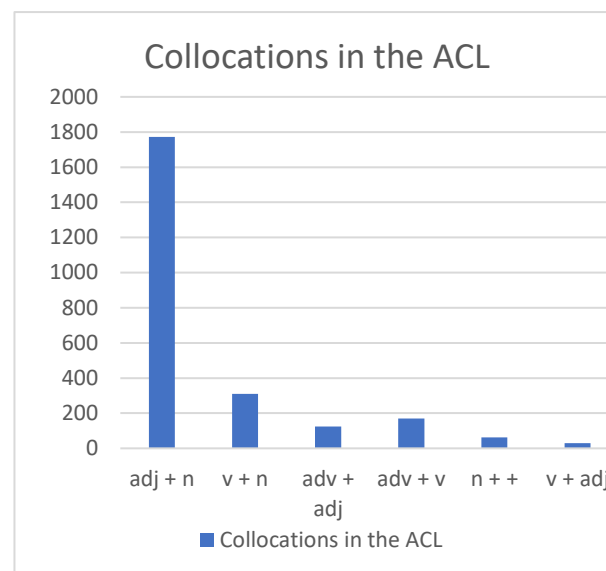


Figure 7.6 Distribution of categories of collocations in the ACL

categories in the ACL and the NACL are visually illustrated in Figure 7.5 and 7.6 respectively, highlighting the contrasting distribution patterns.

The fourth improvement of the New Academic Collocation List is that it was compiled based on a relatively recent and up-to-date database, namely LOCRA, which reflects academic language usage from approximately 2010 to 2022. In contrast, the Academic Collocation List (ACL) was compiled from the Pearson International Corpus of Academic English, which contains data collected over a wider time span, including some portions dating back to 1956. By utilizing a recent corpus for compiling the NACL, it ensures a more accurate representation of the language as it is currently used in academic contexts. The use of a more recent corpus allows for the identification of contemporary phraseological units, reflecting the most up-to-date language patterns and collocations in academic English. This aspect of the NACL enhances its validity and relevance in examining English EFL learners' proficiency in their second language acquisition.

Among the New Academic Collocation List and the Academic Collocation List, 737 collocations appear in both lists, representing approximately 30% of the total size of the ACL.

These shared collocations are distributed across different syntactic dependencies. Out of the 737 shared collocations, 403 are collocations of adjectival modifiers, which account for approximately 54.7% of all shared collocations. This indicates a significant overlap in *amod* collocations between the two lists. In terms of *dobj* collocations, there are 131 shared collocations, representing approximately 17.8% of all the shared collocations. This suggests a considerable intersection in *dobj* collocations between the NACL and the ACL. Furthermore, there are 170 adverb + verb collocations and 40 adverb + adjective collocations among the shared collocations. These account for approximately 23.1% and 4.5% of all shared collocations, respectively. This indicates that there is also a notable overlap in adverbial modifier collocations, specifically adverb + verb and adverb + adjective combinations, between the two lists.

A closer examination of three groups of collocation (i.e., collocations shared by both the NACL and the ACL, collocations exclusively in the NACL, and collocations exclusively in the ACL) yields some interesting results. Firstly, the mean relative frequency for the shared collocations by the ACL and the NACL (i.e., 7.34 per million tokens) exceeds the mean relative frequency of the collocations exclusively in the NACL without the shared collocations (i.e., 5.31 per million tokens). This shows that the collocations shared by the two lists have a higher relative frequency than the collocations that are exclusively included in the NACL. Some of the shared collocations by the NACL and the ACL with the highest relative frequencies are high level, low level, play role, significant effect, and statistically significant. Additionally, the mean mutual information scores for the shared collocations by the ACL and the NACL (i.e., 2.58) exceeds the mean relative frequency of the collocations exclusively in the NACL without the shared collocations (i.e., 1.49). A larger mutual information score shows that the elements of the shared collocations by the two lists are more strongly associated with each other. Some of the highest shared collocations with the highest mutual information scores are informed consent, stark

contrast, vast majority, cast doubt, notable exception, sharp contrast. Table 7.5 and 7.6 lists some additional shared collocations by the two lists and collocations exclusively included in the NACL.

<i>amod</i>	<i>advmod</i>	<i>dobj</i>
<i>active role, alternative method, alternative strategy, available information, careful attention, cultural identity, general trend, high level, main feature, negative impact, political context, primary reason, significant number, traditional practice</i>	<i>equally apply, differently behave, significantly differ, dramatically increase, frequently occur, effectively use, greatly vary, highly significant, mutually exclusive, particularly relevant, qualitatively different, readily available</i>	<i>acquire knowledge, affect outcome, assume role, cast doubt, close relationship, collect information, conduct study, have potential, identify area, make statement, provide summary, seek information, stark contrast, take approach, use resource, use theory</i>

Table 7.5 Examples of the collocations shared by the NACL and the ACL

<i>amod</i>	<i>advmod</i>	<i>dobj</i>
<i>long term, same time, short term, large number, different way, total number, different value, valuable insight, possible interpretation, potential role, current theory, horizontal line, great concern, possible solution</i>	<i>exclusively focus, similarly find, widely accept, simply put, specifically focus, moreover show, indeed find, finally show, first present, however demonstrate, mainly focus, significantly influence, thus serve</i>	<i>perform task, find support, examine relationship, show difference, receive support, meet need, fill gap, mediate relationship, produce effect, occupy position, reach level, make change, explore effect, discuss implication, reduce number</i>

Table 7.6 Examples of the collocations exclusively included in the NACL

A closer comparison between the collocations exclusively included in the ACL and the collocations exclusively included in the NACL reveals that the NACL contains a number of collocations that are specifically associated with the domain of social sciences and humanities. Such collocations include, for instance, *social science, human resource, social process, psychological factor, social behavior*. The collocations included exclusively in the ACL, in contrast, do not include such collocations that are specifically associated with a certain domain. This is probably due to the objectives of these two lists and the corpora used to compile these two lists are different. The NACL is designed to examine L2 EFL learners' writings in the domain of social sciences and humanities. The ACL, on the other hand, is designed as pedagogical resources

for EAP teachers and students. Additionally, the corpus used to compile the NACL (i.e., the nine selected disciplines of the LOCRA) consists of disciplines mainly on social sciences and humanities. The corpus used to compile the ACL (i.e., Pearson International Corpus of Academic English) consists of a wide range of disciplines, not limited to social sciences and humanities.

Table 7.7 shows some examples of the collocations exclusively included in the ACL.

<i>amod</i>	<i>advmod</i>	<i>dobj</i>
<i>academic career, academic performance, accurate assessment, annual report, alternative solution, compelling argument, considerable research, electronic access, individual item, legal status, natural law, professional support,</i>	<i>greatly increase, highly correlate, historically specific, ideally suit, naturally occurring, newly emerging, particularly specific, socially responsible, equally treat, fully realize, greatly enhance</i>	<i>accept responsibility, allow access, attend conference, change attitude, commit crime, conduct survey, deny access, enhance learning, face discrimination, follow instruction, fulfil obligation, give feedback, give priority, have strategy</i>

Table 7.7 Examples of the collocations exclusively included in the ACL

In summary, the shared collocations by the two lists exhibit a higher frequency and a higher degree of association. Additionally, compared to the ACL, the NACL contains more collocations specifically associated with social sciences and humanities, thus is more capable of capturing the most frequent collocations in this domain exclusively. The ACL, on the other hand, contains collocations that are common among a wider range of academic disciplines. This shows that the NACL is more suitable for examining L2 EFL writings of social sciences and humanities whereas the ACL is more suitable as pedagogical resources for EAP teachers and students of various disciplines.

Application of the NACL

The development of an effective and comprehensive academic collocation list is a significant endeavor that holds immense potential for examining L2 writings of EFL learners. This New Academic Collocation List, which we have completed, represents a carefully curated compilation of collocations specific to academic English. It aims to provide language testers and

learner corpus researchers with a valuable linguistic resource to navigate the intricacies of academic discourse and examine L2 writings of EFL learners.

In this phase of the discussion section, we turn our attention to a critical aspect of the academic collocation list's validation and utility – its examination through the lens of learner texts. The analysis of learner texts, produced by individuals at varying levels of language proficiency, allows us to gauge the applicability, effectiveness, and relevance of the collocation list in real-world language use. By observing how learners incorporate these collocations into their writing, we gain insights into the list's impact on language production, as well as potential areas for refinement and improvement.

The application of the NACL section aims to replicate the ACL-derived phraseological sophistication in Paquot (2019) and is designed as follows. A number of learner texts which are written by learners of different proficiency levels are collected. Then, these texts are analyzed by the author manually and three types of collocations are extracted, namely, adjectival modifiers (adj + noun), adverbial modifiers (adv + adj and adv + verb), and direct objects (verb + noun). Then, the New Academic Collocation List is referred to, and the collocations produced by the learners are searched and keep note on which collocations are included in the NACL. Finally, statistical tests are conducted. This replication study has the following hypothesis:

Hypothesis: As learners progress to higher proficiency levels, they produce a higher proportion of potential collocations that are included in the New Academic Collocation List.

To apply the NACL, the Varieties of English for Specific Purposes dAtabase (VESPA; Paquot et al, 2022) is used. The VESPA learner corpus is a learner corpus consists of academic proposals, reports and research papers written by learners of English from a variety of first language backgrounds. To replicate the phraseological sophistication study in Paquot (2019), it is crucial to collect learner texts written by different proficiency levels. A closer examination of the

VESPA corpus shows that the learner texts in the VESPA did not receive a systematic proficiency rating based on Common European Framework of Reference. This means that no learner texts based on their CEFR levels such as B1, B2 and C1 are available to be collected. However, there are some other variables that can be used to function as “pseudo proficiency levels”. The VESPA corpus incorporated the variable of time of learning English at university. Ideally, the learner texts should be rated by professional raters so as to assign accurate proficiency levels. However, no such proficiency level information is provided, and only the variable of years of learning English at university can be used to simulate proficiency information.

In addition to the variable of the time of studying English in university, this study also controls the disciplines of the learner texts to linguistics. The L1 of the learners are controlled for French. The learner text type is controlled as research paper. Two variables of the time spent in university to study English are available, namely two years and six years. No additional variables are controlled. Considering the feasibility of the manual analysis, eight research papers written by French EFL learners are selected. This results in two groups of linguistic research papers written by French EFL learners, 4 of which are written by students who studied English for 2 years in university and 4 of which are written by students who studied English for 6 years in university. The group of texts which are written by students who studied English in university for 2 years are marked as group A. The group of texts that are written by students who studied English in university for 6 years are marked as group B.

The next step involves manually extracting the three types of potential collocations (i.e., *amod*, *advmod*, and *dobj*) from the eight learner texts. Figure 7.7 and 7.8 shows the distribution of potential collocations produced by learners who studied English in university for 2 years and 6 years respectively. The x-axis represents different learners and different types of potential collocations. The y-axis represents the number of each type of potential collocation. It is obvious that the *amod* potential collocations account for the biggest proportion of potential collocations. The *advmod* and *dobj* potential collocations account for a similar proportion. For texts written by learners who studied English in university for 2 years, the average number of *amod*, *advmod*, and *dobj* potential collocations is 109, 44, and 34 respectively. For texts written by learners who studied English in university for 6 years, the average number of *amod*, *advmod*, and *dobj* potential collocations is 63, 25, and 31 respectively. It is obvious that learners who studied English for 6 years in university produced a smaller number of collocations of adjectival modifiers and adverbial modifiers.

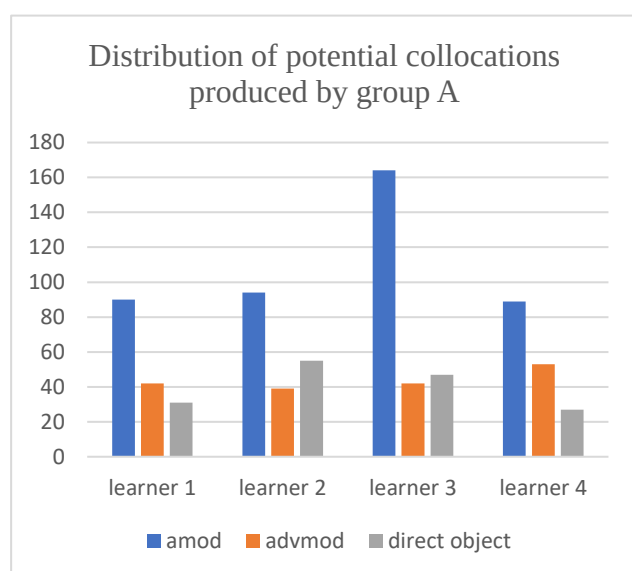


Figure 7.7 Frequency distribution of the three types of potential collocations produced by learners from group A

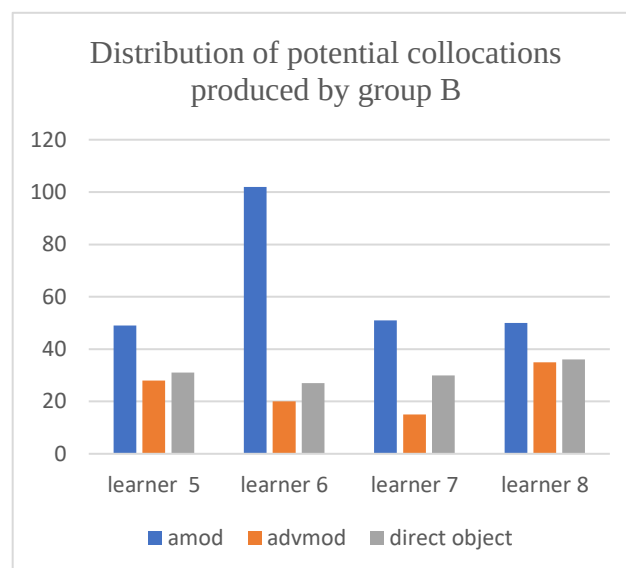


Figure 7.8 Frequency distribution of the three types of potential collocations produced by learners from group B

After manually extracting the collocations existing in the 8 learner texts, it is not to identify which collocations are “sophisticated”, that is, which collocations appear in the New Academic Collocation List. Then, statistical tests are conducted based on these data.

Result

	Mean Group A	Mean Group B	Between-group comparison
<i>amod</i>	0.08	0.15	p = 0.003
<i>advmod</i>	0.14	0.20	p = 0.09
<i>dobj</i>	0.15	0.19	p = 0.42

Table 7.8 Measures of phraseological sophistication based on the NACL

Table 7.8 provides mean percentages of *amod*, *advmod*, and *dobj* potential collocations that are included in the NACL for groups A and B. It can be observed that there is systematic increase of the portion of the “sophisticated” collocations produced from group A to B for the three types of collocations. The statistical tests Chi-square tests were conducted (the critical p value was set to 0.05) and the result showed that only the increase of *amod* potential collocations from group A to B was statistically significant. The increase of *advmod* and *dobj* potential collocations from group A to B is not statistically significant.

The systematic increase shows that as learners’ English competence grows (assuming that a longer time studying English in university guarantees better competence), they produce a higher proportion of collocations that are included in the New Academic Collocation List. This pattern underscores the NACL's efficacy in encapsulating the incremental enhancement of learners' linguistic competence, particularly within the realm of phraseological sophistication.

However, due to the limited availability of learner corpus resources, this application analysis is extremely limited. First, the learner texts should receive professional proficiency ratings, instead of using other variables such as the time spent to study English in university to simulate proficiency levels. Second, the sample size is relatively small. In this analysis, only

eight learner texts were analyzed. This is a rather compromise of available learner corpus resources, which lacks validity. In future research to apply the NACL, a larger sample size should be analyzed. Additionally, these learner texts should receive professional ratings so as to investigate the portion of sophisticated collocations among various proficiency levels.

It is worth noting that during the extraction of potential collocations of the three types of syntactic dependencies in the manual analysis process, some potential collocations repeated abundantly. These potential collocations are closely related to the topic of the research. For instance, in a learner text written by a student who studied English for 2 years in the university, the *amod* potential collocation *neoclassical compound*, appeared 52 times, accounting for 32% of the total *amod* potential collocations (no. 162). In this text, the learner discussed the use of neoclassical compound (e.g., *musicology*) in scientific texts. In this case, it is inevitable that the learner refers to this phenomenon a lot. Some other potential collocations that repeat abundantly in other learner texts are as follows:

backward compatibility, semantic change, new word, bilingual dictionary

All these potential collocations listed above account for more than 10% of all the potential collocations of the adjectival modifier dependency. These repeatedly occurring potential collocations might attenuate the portion of the NACL collocations among all the potential collocations produced by learners, thus causing inaccuracy in the result. This might also explain the statistically non-significant systematic increase as resulted in this learner text study. For future research, researchers should pay additional attention to these potential collocations and propose a new way of tackling such potential collocations in the calculation of phraseological sophistication.

Limitation

While the NACL has made significant improvements, there are still limitations the NACL. One aspect is that the coverage comparison of the NACL and the ACL is based on the LOCRA, where the NACL is extracted from. To compare the coverage of these two lists more objectively, more academic corpora such as the Corpus of Journal Articles (Kosem, 2010) should be incorporated. This requires that the coverage of the NACL and the ACL should be calculated in these academic corpora and compare the difference of coverage. By doing so, a more objective coverage comparison of these two lists is acquired.

Another limitation is that the application of the NACL into learner texts carried out in this research is extremely limited. First, the sample size is relatively small: only eight learner texts are analyzed. Compared to other research such as Paquot (2019) where a total of 98 learner texts were used, eight learner texts were definitely too small. The syntactic dependencies in learner texts were extracted manually, which is extremely time consuming. For future research, professional NLP tools such as spaCy should be incorporated to extract potential collocations. Additionally, the learner texts in this research did not receive professional ratings. Instead, the time of studying English in university is used to categorize low and high proficiency levels, which lacks accuracy. For future research, the learner texts used in the study of phraseological sophistication should receive professional ratings based on CEFR.

Furthermore, the keyness of the collocations in the NACL was established by comparing their relative frequencies in the source corpus (LOCRA) with their corresponding frequencies in a reference non-academic corpus (specifically, the non-academic sections of the BNC were used). While this approach ensures a measure of keyness, there is still potential for improvement in assessing collocation keyness. One possible enhancement is the utilization of the log-likelihood ratio, as demonstrated by Durrant (2009). The log-likelihood ratio is a statistical measure that compares the observed frequency of a collocation in the source corpus with the expected

frequency based on chance alone. By incorporating statistical measures like the log-likelihood ratio, the determination of collocation keyness in the NACL can be further refined. This would provide a more precise and reliable indication of the extent to which a collocation is associated with academic language. Exploring the application of such statistical measures is a valuable direction for future research to enhance the assessment of collocation keyness in the NACL.

8. Conclusion

In conclusion, this research dedicates to the development of a new academic collocation list for research purposes, especially for examining the phraseological competence of L2 EFL learners based on their writings and applying in the studies of phraseological sophistication (Paquot, 2019). Extracted from the selected 9 disciplines of Louvain Corpus of Research Articles which contains approximately 18 million tokens covering a wide range of disciplines of social sciences and humanities, the new list focuses on three syntactic dependencies (i.e., adjectival modifiers, adverbial modifiers, and direct object) and the constructs of frequency and dispersion due to a lack of measuring phraseological sophistication by these constructs. Finally, the New Academic Collocation List (NACL) contains 3,756 collocations (1,497 *amod* collocations, 1361 *advmod* collocations, and 898 *dobj* collocations).

An examination of the NACL reveals that compared to the Academic Collocation List (ACL; Ackermann & Chen, 2013), the NACL contains more collocations than the ACL, which contains 2,496 collocations, providing a more extensive inventory of phraseological resources representative of academic language. Additionally, the NACL contains two times more coverage than the ACL in the LOCRA as calculated in Chapter 7. With its larger coverage from various aspects, the NACL becomes a more representative and comprehensive phraseological resource for assessing English EFL learners' writing. It captures a greater proportion of collocations used in academic contexts, providing researchers and educators with a more extensive and reliable tool to analyze and evaluate the phraseological sophistication of learners' written language. This accomplishes the primary goal delineated at the outset of this research endeavor.

Although the NACL exhibits numerous strengths, there are areas for further improvement. First, the coverage comparison of the NACL and the ACL is based on the LOCRA, where the

NACL is extracted from. The coverage of the NACL and the ACL should be computed for other academic corpora (e.g., Corpus of Academic Journal Articles). Additionally, the application of the NACL into learner texts in this research is extremely limited due to the small sample size and the way of categorizing learners of different proficiency levels (in this research the time of studying English in university is used to indicate low and high proficiency levels). In the future research, a carefully designed study for applying the NACL into learner texts investigating phraseological sophistication should be carried out.

Overall, the New Academic Collocation List serves as a valuable tool for assessing phraseological sophistication in academic English and supports the advancement of language research and pedagogy in this domain. Its comprehensive coverage, alignment with current language usage make it an indispensable resource for researchers, seeking to examine phraseological sophistication of EFL writings.

9. References

- Ackermann, K., & Chen, Y. (2013). Developing the academic collocation list (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235-247.
- Alexander, R. (1984). Fixed expressions in English: Reference books and the teacher. *ELT Journal*, 38(2), 127–134.
- Arbor, A. (2009). Michigan Corpus of Upper-level Student Papers (MICUSP).
<https://quod.lib.umich.edu/m/micusp/>.
- Barkema, H. (1996). Idiomatic and terminology: A multi-dimensional descriptive model. *Studia Linguistica*, 50(2), 125-160.
- Benson, M., Benson, E., & Ilson, R. (1986). *The BBIDictionary of English word combinations*. Amsterdam: Benjamins.
- Bestgen, Y., & Granger, S. (2018). Tracking L2 writers' phraseological development using Collgrams: Evidence from a longitudinal EFL corpus. In H. Sebastian , A. Sand, A. Sabine, & L. Dillmann, *Corpora and lexis* (pp. 277–301). Brill: Brill Rodopi.
- Chen, Y., & Baker, P. (2010). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6), 849–880.
- Chon, Y., & Shin, D. (2013). A corpus-driven analysis of spoken and written academic. *Multimedia-Assisted Language Learning*, 16, 11-38.

- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning*. Cambridge: Cambridge University Press.
- Cowie, A. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223–235.
- Cowie, A. (1994). Phraseology. In R. Asher, *The Encyclopaedia of Language and Linguistics* (pp. 3168-3171). Oxford: Pergamon.
- Cowie, A. (1997). Phraseology in formal academic prose. In J. Aarts, I. De Mönnink, & H. Wekker, *Studies in English language and teaching : in honor of Flor Aarts* (pp. 43-56). Amsterdam and Atlanta: Rodopi.
- Cowie, A. (1998). *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press.
- Cowie, A. (2001). Exploring native-speaker knowledge of phraseology: informant testing or corpus research? In H. Burger, A. Buhofer, & G. Greciano, *Flut von Texten -Vielfalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifität der Phraseologie* (pp. 73-81). Essen: Schmeider Verlag Hogengerehen GmbH.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier, & S. Granger, *Phraseology in foreign language learning and teaching* (pp. 149-161). Amsterdam/Philadelphia: John Benjamins.
- Davies, M. (2004). British National Corpus (from Oxford University Press). Available online at <https://www.english-corpora.org/bnc/>.

- Davies, M. (2008-). The Corpus of Contemporary American English (COCA). Available online at <https://www.english-corpora.org/coca/>.
- De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. *Corpus Linguistics and Linguistic Theory*, 51-68.
- De Cock, S. (2003). *Recurrent sequences of words in native speaker and advanced learner spoken and written English: a corpus-driven approach*. Louvain-la-Neuve: Universite catholique de Louvain: Unpublished PhD thesis.
- De Marneffe, M.-C., & Manning, C. (2017, Februaray). *Stanford typed dependencies manual*. Retrieved from http://nlp.stanford.edu/software/dependenceis_manual.pdf
- Durrant , P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47(2), 157-177.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Academic Purposes*, 28(3), 157-169.
- Gardner, D., & Davies, M. (2014). New Academic Vocabulary List. *Applied Linguistics*, 305–327.
- Gramley, S., & Patzold, M. (1992). *A survey of Modern English*. London: Routledge.
- Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson, *Languages in Contrast. Text-based cross-linguistic studies* (pp. 37–51). Lund: Lund University Press.

- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. Cowie, *Phraseology: theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, 52(3), 229-252.
- Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (2020). The International Corpus of Learner English. Version 3. Louvain-la-Neuve: Presses universitaires de Louvain.
- Gries, S. T. (2008). Disentangling the phraseological web. In S. Granger , & F. Meunier , *Phraseology: An interdisciplinary perspective* (pp. 27–49). Amsterdam : John Benjamins.
- Gross, G. (1996). *Les expressions figées en français: noms composés et autres locutions*. Paris: Éditions Ophrys.
- Halliday, M. (1961). *Categories of the theory of grammar*. Word 17.
- Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Howarth, P. (1996). *Phraseology in English Academic Writing: Some Implications for language learning and dictionary making*. Tübingen: Max Niemeyer Verlag.
- Hsu, W. (2014). The most frequent opaque formulaic sequences in English-medium college textbooks. *System*, 47, 146-161.
- Jones, S., & Sinclair, J. (1974). English lexical collocations. *Cahiers de Lexicologie*, 24, 15-61.
- Kosem, I. (2010). Designing a model for a corpus-driven dictionary of Academic English. *PhD Thesis*. Aston University.

- Lee, S. (2006). A corpus-based analysis of Korean EFL learner's use of amplifier collocations. *English Teaching*, 61(1), 3-17.
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35.
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299–320.
- Mel'čuk, I. (1998). Collocations and lexical functions. In A. Cowie, *Phraseology: Theory, analysis, and applications* (pp. 23-53). Oxford: Oxford University Press.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Paquot, M. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London & New-York: Continuum.
- Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly*, 15(1), 29-43.
- Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second language research*, 35(1), 121-145.
- Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130-149.
- Paquot, M., Larsson, T., Hasselgård, H., Ebeling, S., De Meyere, D., Valentin, L., . . . van Vuuren, S. (2022). The Varieties of English for Specific Purposes dAtabase (VESPA): Towards a

multi-L1 and multi-register learner corpus of disciplinary writing. *Research in Corpus Linguistics*, 10(2), 1-5.

Rogers, J., Müller, A., Daulton, F., Dickinson, P., Florescu, C., Reid, G., & Stoeckel, T. (2021). The creation and application of a large-scale corpus-based academic multi-word unit list. *English for Specific Purposes*, 62, 142-157.

Rubin, R., Housen, A., & Paquot, M. (2021). Phraseological complexity as an index of L2 Dutch writing proficiency: A partial replication study. In S. Granger, *Perspectives on the L2 Phrasicon: The View from Learner Corpora* (pp. 101-125). Bristol: Multilingual Matters.

Shin, D., & Nation, P. (2008). Beyond single words: the most frequent collocations in spoken. *ELT Journal*, 62(4), 339–348.

Simpson-Vlach, R., & Ellis, R. (2010). An academic formulas list: new methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429-458.

Swales, R., Briggs, J., Ovens, S., & Simpson, J. (2007). Michigan Corpus of Academic Spoken English. Available at <http://quod.lib.umich.edu/m/micase/>.

Van Roey, J. (1990). *French-English contrastive lexicology: An introduction*. Leuven: Peeters.

- Vandeweerd, N., Housen, A., & Paquot, M. (2021). Applying phraseological complexity measures to L2 French: A partial replication study. *International Journal of Learner Corpus Research*, 7(2), 197-229.
- Vandeweerd, N., Housen, A., & Paquot, M. (2022). Comparing the longitudinal development of phraseological complexity across oral and written tasks. *Studies in Second Language Acquisition*, 1 - 25.
- Vinogradov, V. (1953). On some issues of Russian historical lexicology. *Bulletin of the USSR Academy of Sciences, Language and Literature Section*, 12(3), 185-210.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

10. Appendix

The appendix lists all the entries in the New Academic Collocation List.

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1	high	adj	level	noun	40	social	adj	network	noun
2	long	adj	term	noun	41	third	adj	party	noun
3	other	adj	word	noun	42	recent	adj	year	noun
4	other	adj	hand	noun	43	other	adj	group	noun
5	same	adj	time	noun	44	empirical	adj	evidence	noun
6	present	adj	study	noun	45	economic	adj	development	noun
7	human	adj	right	noun	46	large	adj	scale	noun
8	low	adj	level	noun	47	first	adj	time	noun
9	significant	adj	effect	noun	48	civil	adj	war	noun
10	significant	adj	difference	noun	49	social	adj	interaction	noun
11	previous	adj	study	noun	50	empirical	adj	study	noun
12	previous	adj	research	noun	51	full	adj	time	noun
13	current	adj	study	noun	52	other	adj	form	noun
14	short	adj	term	noun	53	social	adj	context	noun
15	high	adj	school	noun	54	direct	adj	effect	noun
16	middle	adj	class	noun	55	national	adj	level	noun
17	different	adj	type	noun	56	total	adj	number	noun
18	political	adj	party	noun	57	public	adj	sector	noun
19	other	adj	country	noun	58	high	adj	quality	noun
20	large	adj	number	noun	59	further	adj	research	noun
21	social	adj	capital	noun	60	social	adj	science	noun
22	important	adj	role	noun	61	next	adj	section	noun
23	negative	adj	effect	noun	62	same	adj	way	noun
24	different	adj	level	noun	63	small	adj	number	noun
25	wide	adj	range	noun	64	early	adj	stage	noun
26	public	adj	policy	noun	65	several	adj	study	noun
27	other	adj	study	noun	66	theoretical	adj	framework	noun
28	other	adj	factor	noun	67	high	adj	degree	noun
29	young	adj	people	noun	68	few	adj	study	noun
30	high	adj	education	noun	69	past	adj	decade	noun
31	high	adj	rate	noun	70	English	adj	language	noun
32	first	adj	year	noun	71	sexual	adj	orientation	noun
33	social	adj	identity	noun	72	other	adj	member	noun
34	recent	adj	study	noun	73	public	adj	school	noun
35	different	adj	way	noun	74	public	adj	opinion	noun
36	other	adj	variable	noun	75	social	adj	relation	noun
37	ethnic	adj	group	noun	76	different	adj	group	noun
38	future	adj	study	noun	77	high	adj	value	noun
39	foreign	adj	language	noun	78	human	adj	being	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
79	second	adj	order	noun	123	significant	adj	relationship	noun
80	internal	adj	consistency	noun	124	long	adj	period	noun
81	everyday	adj	life	noun	125	previous	adj	work	noun
82	social	adj	change	noun	126	vast	adj	majority	noun
83	cognitive	adj	process	noun	127	social	adj	structure	noun
84	secondary	adj	school	noun	128	more	adj	time	noun
85	social	adj	life	noun	129	such	adj	case	noun
86	first	adj	step	noun	130	good	adj	interest	noun
87	private	adj	sector	noun	131	statistical	adj	significance	noun
88	small	adj	group	noun	132	new	adj	technology	noun
89	other	adj	type	noun	133	other	adj	area	noun
90	cultural	adj	value	noun	134	many	adj	people	noun
91	last	adj	year	noun	135	high	adj	status	noun
92	past	adj	year	noun	136	negative	adj	impact	noun
93	different	adj	context	noun	137	different	adj	kind	noun
94	many	adj	study	noun	138	small	adj	business	noun
95	social	adj	group	noun	139	possible	adj	explanation	noun
96	social	adj	class	noun	140	similar	adj	result	noun
97	recent	adj	research	noun	141	local	adj	community	noun
98	important	adj	implication	noun	142	other	adj	case	noun
99	ethnic	adj	minority	noun	143	well	adj	understanding	noun
100	many	adj	country	noun	144	social	adj	norm	noun
101	empirical	adj	research	noun	145	high	adj	proportion	noun
102	first	adj	place	noun	146	other	adj	thing	noun
103	different	adj	form	noun	147	local	adj	level	noun
104	last	adj	decade	noun	148	similar	adj	pattern	noun
105	new	adj	form	noun	149	strong	adj	relationship	noun
106	low	adj	rate	noun	150	urban	adj	area	noun
107	high	adj	order	noun	151	other	adj	aspect	noun
108	national	adj	identity	noun	152	social	adj	relationship	noun
109	many	adj	case	noun	153	competitive	adj	advantage	noun
110	statistical	adj	analysis	noun	154	other	adj	context	noun
111	real	adj	world	noun	155	other	adj	actor	noun
112	great	adj	deal	noun	156	political	adj	context	noun
113	important	adj	factor	noun	157	significant	adj	change	noun
114	more	adj	detail	noun	158	english	adj	speaker	noun
115	other	adj	language	noun	159	rural	adj	area	noun
116	recent	adj	work	noun	160	more	adj	attention	noun
117	negative	adj	emotion	noun	161	previous	adj	section	noun
118	other	adj	state	noun	162	other	adj	way	noun
119	many	adj	way	noun	163	positive	adj	outcome	noun
120	long	adj	time	noun	164	large	adj	size	noun
121	daily	adj	life	noun	165	alternative	adj	explanation	noun
122	social	adj	medium	noun	166	social	adj	status	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
167	key	adj	role	noun	211	first	adj	author	noun
168	most	adj	case	noun	212	real	adj	time	noun
169	theoretical	adj	model	noun	213	important	adj	question	noun
170	good	adj	practice	noun	214	open	adj	question	noun
171	broad	adj	range	noun	215	further	adj	analysis	noun
172	socioeconomic	adj	status	noun	216	new	adj	information	noun
173	social	adj	order	noun	217	little	adj	evidence	noun
174	high	adj	number	noun	218	informed	adj	consent	noun
175	financial	adj	resource	noun	219	negative	adj	consequence	noun
176	metropolitan	adj	area	noun	220	high	adj	cost	noun
177	different	adj	aspect	noun	221	more	adj	information	noun
178	wide	adj	variety	noun	222	important	adj	aspect	noun
179	new	adj	model	noun	223	other	adj	source	noun
180	long	adj	history	noun	224	relative	adj	importance	noun
181	specific	adj	context	noun	225	old	adj	age	noun
182	new	adj	idea	noun	226	strong	adj	evidence	noun
183	large	adj	group	noun	227	significant	adj	impact	noun
184	many	adj	year	noun	228	raw	adj	material	noun
185	new	adj	way	noun	229	own	adj	right	noun
186	social	adj	justice	noun	230	little	adj	attention	noun
187	social	adj	world	noun	231	public	adj	health	noun
188	large	adj	part	noun	232	same	adj	level	noun
189	different	adj	country	noun	233	social	adj	organization	noun
190	cultural	adj	difference	noun	234	most	adj	study	noun
191	real	adj	life	noun	235	own	adj	experience	noun
192	social	adj	practice	noun	236	social	adj	environment	noun
193	various	adj	form	noun	237	recent	adj	decade	noun
194	cultural	adj	context	noun	238	significant	adj	role	noun
195	other	adj	party	noun	239	first	adj	order	noun
196	potential	adj	effect	noun	240	first	adj	person	noun
197	different	adj	pattern	noun	241	second	adj	half	noun
198	early	adj	study	noun	242	close	adj	relationship	noun
199	first	adj	study	noun	243	human	adj	resource	noun
200	great	adj	number	noun	244	similar	adj	effect	noun
201	central	adj	role	noun	245	other	adj	participant	noun
202	small	adj	scale	noun	246	several	adj	reason	noun
203	most	adj	part	noun	247	financial	adj	support	noun
204	few	adj	year	noun	248	great	adj	extent	noun
205	personal	adj	experience	noun	249	native	adj	language	noun
206	different	adj	approach	noun	250	social	adj	mobility	noun
207	first	adj	stage	noun	251	theoretical	adj	perspective	noun
208	early	adj	year	noun	252	national	adj	policy	noun
209	economic	adj	activity	noun	253	similar	adj	way	noun
210	positive	adj	attitude	noun	254	such	adj	effect	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
255	different	adj	perspective	noun	299	cultural	adj	heritage	noun
256	social	adj	process	noun	300	other	adj	institution	noun
257	social	adj	theory	noun	301	cultural	adj	norm	noun
258	important	adj	difference	noun	302	important	adj	part	noun
259	practical	adj	implication	noun	303	other	adj	scholar	noun
260	other	adj	part	noun	304	several	adj	way	noun
261	low	adj	status	noun	305	such	adj	change	noun
262	various	adj	way	noun	306	cultural	adj	practice	noun
263	previous	adj	year	noun	307	other	adj	region	noun
264	most	adj	people	noun	308	rapid	adj	growth	noun
265	important	adj	issue	noun	309	other	adj	student	noun
266	long	adj	run	noun	310	social	adj	problem	noun
267	human	adj	body	noun	311	second	adj	question	noun
268	current	adj	research	noun	312	causal	adj	relationship	noun
269	available	adj	resource	noun	313	academic	adj	research	noun
270	key	adj	factor	noun	314	several	adj	year	noun
271	first	adj	question	noun	315	same	adj	pattern	noun
272	necessary	adj	condition	noun	316	crucial	adj	role	noun
273	different	adj	language	noun	317	local	adj	context	noun
274	social	adj	reality	noun	318	less	adj	extent	noun
275	racial	adj	group	noun	319	right	adj	hand	noun
276	poor	adj	performance	noun	320	other	adj	model	noun
277	short	adj	period	noun	321	more	adj	resource	noun
278	low	adj	value	noun	322	other	adj	activity	noun
279	young	adj	age	noun	323	good	adj	reason	noun
280	large	adj	difference	noun	324	different	adj	source	noun
281	other	adj	side	noun	325	particular	adj	type	noun
282	deep	adj	understanding	noun	326	other	adj	work	noun
283	social	adj	issue	noun	327	educational	adj	level	noun
284	critical	adj	role	noun	328	such	adj	study	noun
285	early	adj	work	noun	329	social	adj	system	noun
286	particular	adj	context	noun	330	social	adj	value	noun
287	different	adj	dimension	noun	331	important	adj	contribution	noun
288	next	adj	step	noun	332	other	adj	organization	noun
289	present	adj	day	noun	333	small	adj	sample	noun
290	local	adj	knowledge	noun	334	social	adj	service	noun
291	general	adj	principle	noun	335	subsequent	adj	analysis	noun
292	previous	adj	experience	noun	336	alternative	adj	model	noun
293	ethnic	adj	identity	noun	337	such	adj	practice	noun
294	many	adj	scholar	noun	338	broad	adj	context	noun
295	different	adj	stage	noun	339	such	adj	difference	noun
296	high	adj	stake	noun	340	further	adj	evidence	noun
297	new	adj	opportunity	noun	341	new	adj	insight	noun
298	cross	adj	study	noun	342	first	adj	part	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
343	social	adj	actor	noun	387	different	adj	model	noun
344	detailed	adj	analysis	noun	388	general	adj	public	noun
345	other	adj	characteristic	noun	389	other	adj	category	noun
346	limited	adj	number	noun	390	social	adj	category	noun
347	many	adj	other	noun	391	available	adj	information	noun
348	main	adj	reason	noun	392	second	adj	part	noun
349	social	adj	scientist	noun	393	great	adj	power	noun
350	different	adj	set	noun	394	short	adj	time	noun
351	western	adj	country	noun	395	key	adj	aspect	noun
352	primary	adj	education	noun	396	other	adj	dimension	noun
353	important	adj	way	noun	397	different	adj	point	noun
354	only	adj	way	noun	398	social	adj	position	noun
355	first	adj	instance	noun	399	emotional	adj	response	noun
356	first	adj	model	noun	400	possible	adj	effect	noun
357	social	adj	activity	noun	401	past	adj	experience	noun
358	different	adj	area	noun	402	certain	adj	type	noun
359	different	adj	background	noun	403	particular	adj	way	noun
360	key	adj	element	noun	404	adverse	adj	effect	noun
361	large	adj	extent	noun	405	detailed	adj	information	noun
362	theoretical	adj	approach	noun	406	other	adj	effect	noun
363	first	adj	glance	noun	407	several	adj	decade	noun
364	own	adj	interest	noun	408	opposite	adj	direction	noun
365	specific	adj	type	noun	409	common	adj	ground	noun
366	large	adj	amount	noun	410	early	adj	period	noun
367	great	adj	degree	noun	411	limited	adj	resource	noun
368	personal	adj	identity	noun	412	different	adj	category	noun
369	free	adj	market	noun	413	various	adj	level	noun
370	explanatory	adj	power	noun	414	indigenous	adj	community	noun
371	less	adj	attention	noun	415	emotional	adj	experience	noun
372	further	adj	investigation	noun	416	same	adj	period	noun
373	other	adj	time	noun	417	different	adj	system	noun
374	various	adj	type	noun	418	other	adj	person	noun
375	key	adj	feature	noun	419	multiple	adj	source	noun
376	particular	adj	interest	noun	420	political	adj	agenda	noun
377	integral	adj	part	noun	421	social	adj	condition	noun
378	social	adj	space	noun	422	important	adj	insight	noun
379	first	adj	half	noun	423	close	adj	tie	noun
380	new	adj	system	noun	424	several	adj	time	noun
381	second	adj	step	noun	425	similar	adj	vein	noun
382	different	adj	effect	noun	426	great	adj	attention	noun
383	other	adj	issue	noun	427	specific	adj	form	noun
384	such	adj	research	noun	428	dominant	adj	group	noun
385	several	adj	factor	noun	429	particular	adj	group	noun
386	potential	adj	benefit	noun	430	cultural	adj	identity	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
431	different	adj	method	noun	475	such	adj	behavior	noun
432	large	adj	population	noun	476	multiple	adj	level	noun
433	different	adj	time	noun	477	key	adj	point	noun
434	great	adj	level	noun	478	complex	adj	relationship	noun
435	focal	adj	point	noun	479	same	adj	thing	noun
436	high	adj	profile	noun	480	same	adj	type	noun
437	complex	adj	system	noun	481	certain	adj	kind	noun
438	small	adj	town	noun	482	new	adj	one	noun
439	social	adj	action	noun	483	major	adj	role	noun
440	broad	adj	category	noun	484	different	adj	setting	noun
441	few	adj	decade	noun	485	specific	adj	aspect	noun
442	total	adj	population	noun	486	close	adj	attention	noun
443	various	adj	aspect	noun	487	external	adj	factor	noun
444	key	adj	issue	noun	488	significant	adj	influence	noun
445	different	adj	position	noun	489	own	adj	practice	noun
446	common	adj	sense	noun	490	same	adj	set	noun
447	extended	adj	period	noun	491	potential	adj	source	noun
448	other	adj	approach	noun	492	specific	adj	area	noun
449	different	adj	process	noun	493	global	adj	market	noun
450	much	adj	attention	noun	494	current	adj	practice	noun
451	formal	adj	education	noun	495	further	adj	support	noun
452	many	adj	form	noun	496	current	adj	context	noun
453	social	adj	integration	noun	497	political	adj	discourse	noun
454	great	adj	value	noun	498	various	adj	factor	noun
455	small	adj	difference	noun	499	certain	adj	group	noun
456	social	adj	factor	noun	500	such	adj	information	noun
457	such	adj	knowledge	noun	501	general	adj	level	noun
458	current	adj	situation	noun	502	new	adj	perspective	noun
459	economic	adj	status	noun	503	large	adj	project	noun
460	specific	adj	way	noun	504	natural	adj	science	noun
461	early	adj	age	noun	505	such	adj	strategy	noun
462	rapid	adj	change	noun	506	different	adj	value	noun
463	same	adj	effect	noun	507	different	adj	view	noun
464	different	adj	part	noun	508	similar	adj	approach	noun
465	specific	adj	practice	noun	509	key	adj	component	noun
466	strong	adj	sense	noun	510	great	adj	detail	noun
467	stark	adj	contrast	noun	511	important	adj	feature	noun
468	social	adj	distance	noun	512	primary	adj	focus	noun
469	racial	adj	diversity	noun	513	specific	adj	group	noun
470	specific	adj	feature	noun	514	social	adj	meaning	noun
471	same	adj	group	noun	515	close	adj	proximity	noun
472	complex	adj	process	noun	516	active	adj	role	noun
473	interesting	adj	question	noun	517	certain	adj	way	noun
474	daily	adj	basis	noun	518	many	adj	aspect	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
519	actual	adj	behavior	noun	563	great	adj	likelihood	noun
520	few	adj	exception	noun	564	very	adj	nature	noun
521	human	adj	experience	noun	565	such	adj	relationship	noun
522	next	adj	generation	noun	566	effective	adj	way	noun
523	diverse	adj	group	noun	567	single	adj	case	noun
524	multiple	adj	identity	noun	568	plausible	adj	explanation	noun
525	individual	adj	interest	noun	569	other	adj	half	noun
526	new	adj	kind	noun	570	first	adj	phase	noun
527	own	adj	position	noun	571	important	adj	step	noun
528	personal	adj	relationship	noun	572	multiple	adj	way	noun
529	low	adj	wage	noun	573	human	adj	agency	noun
530	recent	adj	analysis	noun	574	few	adj	case	noun
531	different	adj	culture	noun	575	same	adj	direction	noun
532	clear	adj	evidence	noun	576	specific	adj	goal	noun
533	second	adj	model	noun	577	great	adj	interest	noun
534	prominent	adj	role	noun	578	own	adj	life	noun
535	alternative	adj	approach	noun	579	such	adj	situation	noun
536	similar	adj	level	noun	580	traditional	adj	approach	noun
537	fundamental	adj	principle	noun	581	subject	adj	area	noun
538	potential	adj	problem	noun	582	positive	adj	change	noun
539	certain	adj	aspect	noun	583	new	adj	environment	noun
540	special	adj	attention	noun	584	third	adj	generation	noun
541	such	adj	circumstance	noun	585	certain	adj	level	noun
542	specific	adj	condition	noun	586	contemporary	adj	society	noun
543	other	adj	field	noun	587	new	adj	state	noun
544	more	adj	half	noun	588	early	adj	childhood	noun
545	strong	adj	influence	noun	589	popular	adj	culture	noun
546	many	adj	instance	noun	590	only	adj	difference	noun
547	strong	adj	link	noun	591	such	adj	effort	noun
548	own	adj	work	noun	592	great	adj	emphasis	noun
549	previous	adj	analysis	noun	593	additional	adj	evidence	noun
550	further	adj	development	noun	594	significant	adj	factor	noun
551	other	adj	end	noun	595	specific	adj	information	noun
552	direct	adj	relationship	noun	596	common	adj	interest	noun
553	western	adj	society	noun	597	various	adj	kind	noun
554	specific	adj	case	noun	598	other	adj	process	noun
555	ongoing	adj	debate	noun	599	recent	adj	scholarship	noun
556	great	adj	effect	noun	600	political	adj	structure	noun
557	recent	adj	history	noun	601	other	adj	system	noun
558	such	adj	model	noun	602	small	adj	amount	noun
559	different	adj	order	noun	603	other	adj	example	noun
560	such	adj	condition	noun	604	such	adj	factor	noun
561	available	adj	evidence	noun	605	direct	adj	impact	noun
562	global	adj	level	noun	606	such	adj	measure	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
607	new	adj	method	noun	651	general	adj	term	noun
608	first	adj	month	noun	652	multiple	adj	time	noun
609	such	adj	question	noun	653	extreme	adj	case	noun
610	different	adj	region	noun	654	wide	adj	context	noun
611	different	adj	role	noun	655	early	adj	day	noun
612	social	adj	role	noun	656	other	adj	difference	noun
613	general	adj	trend	noun	657	important	adj	dimension	noun
614	such	adj	analysis	noun	658	new	adj	direction	noun
615	key	adj	concept	noun	659	low	adj	end	noun
616	individual	adj	identity	noun	660	main	adj	focus	noun
617	own	adj	identity	noun	661	upper	adj	level	noun
618	social	adj	inclusion	noun	662	global	adj	scale	noun
619	specific	adj	issue	noun	663	other	adj	sector	noun
620	same	adj	kind	noun	664	broad	adj	sense	noun
621	social	adj	phenomenon	noun	665	social	adj	stratification	noun
622	serious	adj	problem	noun	666	cultural	adj	tradition	noun
623	specific	adj	situation	noun	667	other	adj	author	noun
624	main	adj	source	noun	668	first	adj	case	noun
625	long	adj	tradition	noun	669	new	adj	concept	noun
626	total	adj	amount	noun	670	social	adj	connection	noun
627	social	adj	behavior	noun	671	sharp	adj	contrast	noun
628	latter	adj	case	noun	672	final	adj	decision	noun
629	past	adj	century	noun	673	important	adj	distinction	noun
630	great	adj	impact	noun	674	future	adj	generation	noun
631	personal	adj	life	noun	675	specific	adj	question	noun
632	ongoing	adj	process	noun	676	considerable	adj	attention	noun
633	reciprocal	adj	relationship	noun	677	specific	adj	behavior	noun
634	frequent	adj	use	noun	678	certain	adj	circumstance	noun
635	significant	adj	variation	noun	679	violent	adj	crime	noun
636	more	adj	child	noun	680	certain	adj	degree	noun
637	other	adj	community	noun	681	social	adj	development	noun
638	good	adj	deal	noun	682	ethical	adj	dilemma	noun
639	detrimental	adj	effect	noun	683	social	adj	dimension	noun
640	notable	adj	exception	noun	684	potential	adj	explanation	noun
641	different	adj	interpretation	noun	685	great	adj	influence	noun
642	recent	adj	literature	noun	686	basic	adj	level	noun
643	traditional	adj	method	noun	687	overwhelming	adj	majority	noun
644	dynamic	adj	model	noun	688	key	adj	mechanism	noun
645	significant	adj	number	noun	689	different	adj	period	noun
646	other	adj	option	noun	690	actual	adj	practice	noun
647	fundamental	adj	question	noun	691	social	adj	pressure	noun
648	main	adj	question	noun	692	different	adj	reason	noun
649	same	adj	result	noun	693	different	adj	result	noun
650	difficult	adj	task	noun	694	internal	adj	structure	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
695	last	adj	century	noun	739	poor	adj	quality	noun
696	such	adj	context	noun	740	high	adj	ratio	noun
697	broad	adj	definition	noun	741	similar	adj	argument	noun
698	economic	adj	factor	noun	742	high	adj	end	noun
699	alternative	adj	form	noun	743	recent	adj	finding	noun
700	valuable	adj	insight	noun	744	second	adj	group	noun
701	more	adj	knowledge	noun	745	complex	adj	interaction	noun
702	same	adj	process	noun	746	few	adj	month	noun
703	interpersonal	adj	relationship	noun	747	few	adj	opportunity	noun
704	new	adj	relationship	noun	748	initial	adj	stage	noun
705	many	adj	respect	noun	749	large	adj	system	noun
706	great	adj	use	noun	750	large	adj	area	noun
707	meaningful	adj	way	noun	751	direct	adj	comparison	noun
708	seminal	adj	work	noun	752	important	adj	consequence	noun
709	significant	adj	amount	noun	753	regional	adj	difference	noun
710	many	adj	area	noun	754	detailed	adj	discussion	noun
711	special	adj	case	noun	755	cultural	adj	diversity	noun
712	such	adj	experience	noun	756	different	adj	environment	noun
713	political	adj	identity	noun	757	various	adj	group	noun
714	different	adj	people	noun	758	theoretical	adj	insight	noun
715	several	adj	scholar	noun	759	possible	adj	interpretation	noun
716	dark	adj	side	noun	760	active	adj	participation	noun
717	possible	adj	way	noun	761	complex	adj	phenomenon	noun
718	significant	adj	contribution	noun	762	basic	adj	principle	noun
719	such	adj	decision	noun	763	different	adj	version	noun
720	negative	adj	experience	noun	764	more	adj	work	noun
721	near	adj	future	noun	765	half	adj	century	noun
722	first	adj	group	noun	766	low	adj	degree	noun
723	further	adj	insight	noun	767	natural	adj	environment	noun
724	such	adj	issue	noun	768	different	adj	experience	noun
725	other	adj	mean	noun	769	common	adj	feature	noun
726	more	adj	money	noun	770	alternative	adj	interpretation	noun
727	great	adj	opportunity	noun	771	great	adj	potential	noun
728	economic	adj	situation	noun	772	same	adj	problem	noun
729	physical	adj	space	noun	773	many	adj	question	noun
730	new	adj	strategy	noun	774	same	adj	reason	noun
731	own	adj	word	noun	775	new	adj	role	noun
732	good	adj	evidence	noun	776	right	adj	side	noun
733	certain	adj	extent	noun	777	major	adj	source	noun
734	distinctive	adj	feature	noun	778	little	adj	support	noun
735	technological	adj	innovation	noun	779	own	adj	term	noun
736	same	adj	manner	noun	780	alternative	adj	way	noun
737	large	adj	network	noun	781	such	adj	approach	noun
738	available	adj	option	noun	782	several	adj	aspect	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
783	several	adj	author	noun	827	theoretical	adj	basis	noun
784	large	adj	majority	noun	828	second	adj	dimension	noun
785	good	adj	model	noun	829	ample	adj	evidence	noun
786	several	adj	month	noun	830	close	adj	examination	noun
787	various	adj	source	noun	831	early	adj	experience	noun
788	first	adj	attempt	noun	832	complex	adj	issue	noun
789	human	adj	condition	noun	833	substantial	adj	number	noun
790	social	adj	contact	noun	834	possible	adj	outcome	noun
791	various	adj	context	noun	835	few	adj	resource	noun
792	large	adj	gap	noun	836	subsequent	adj	study	noun
793	significant	adj	implication	noun	837	initial	adj	support	noun
794	second	adj	level	noun	838	good	adj	thing	noun
795	entire	adj	period	noun	839	many	adj	thing	noun
796	central	adj	question	noun	840	useful	adj	tool	noun
797	fundamental	adj	change	noun	841	new	adj	type	noun
798	different	adj	direction	noun	842	new	adj	understanding	noun
799	other	adj	event	noun	843	main	adj	argument	noun
800	such	adj	form	noun	844	public	adj	attention	noun
801	many	adj	individual	noun	845	modern	adj	day	noun
802	potential	adj	influence	noun	846	further	adj	discussion	noun
803	many	adj	issue	noun	847	compelling	adj	evidence	noun
804	new	adj	question	noun	848	different	adj	function	noun
805	new	adj	resource	noun	849	single	adj	group	noun
806	negative	adj	view	noun	850	first	adj	line	noun
807	much	adj	work	noun	851	overall	adj	pattern	noun
808	final	adj	analysis	noun	852	major	adj	problem	noun
809	other	adj	change	noun	853	similar	adj	process	noun
810	public	adj	domain	noun	854	positive	adj	result	noun
811	clear	adj	example	noun	855	pivotal	adj	role	noun
812	individual	adj	experience	noun	856	different	adj	situation	noun
813	certain	adj	feature	noun	857	appropriate	adj	way	noun
814	new	adj	generation	noun	858	final	adj	year	noun
815	such	adj	interaction	noun	859	methodological	adj	approach	noun
816	new	adj	light	noun	860	very	adj	beginning	noun
817	different	adj	mode	noun	861	small	adj	change	noun
818	new	adj	paradigm	noun	862	everyday	adj	experience	noun
819	entire	adj	population	noun	863	bottom	adj	line	noun
820	traditional	adj	practice	noun	864	same	adj	line	noun
821	other	adj	property	noun	865	other	adj	question	noun
822	various	adj	reason	noun	866	valuable	adj	resource	noun
823	potential	adj	role	noun	867	appropriate	adj	response	noun
824	new	adj	situation	noun	868	own	adj	action	noun
825	same	adj	structure	noun	869	human	adj	activity	noun
826	implicit	adj	assumption	noun	870	individual	adj	agency	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
871	particular	adj	aspect	noun	915	same	adj	context	noun
872	great	adj	benefit	noun	916	great	adj	difference	noun
873	new	adj	challenge	noun	917	new	adj	era	noun
874	several	adj	characteristic	noun	918	critical	adj	issue	noun
875	great	adj	control	noun	919	major	adj	issue	noun
876	major	adj	difference	noun	920	other	adj	problem	noun
877	multiple	adj	factor	noun	921	particular	adj	set	noun
878	salient	adj	feature	noun	922	basic	adj	structure	noun
879	multiple	adj	form	noun	923	different	adj	style	noun
880	similar	adj	manner	noun	924	other	adj	term	noun
881	complex	adj	nature	noun	925	common	adj	understanding	noun
882	specific	adj	policy	noun	926	important	adj	work	noun
883	primary	adj	reason	noun	927	following	adj	year	noun
884	particular	adj	situation	noun	928	cultural	adj	change	noun
885	average	adj	size	noun	929	own	adj	community	noun
886	general	adj	theory	noun	930	similar	adj	conclusion	noun
887	individual	adj	actor	noun	931	male	adj	counterpart	noun
888	certain	adj	area	noun	932	ethnic	adj	difference	noun
889	positive	adj	aspect	noun	933	great	adj	difficulty	noun
890	great	adj	challenge	noun	934	such	adj	evidence	noun
891	certain	adj	characteristic	noun	935	direct	adj	experience	noun
892	strong	adj	connection	noun	936	direct	adj	link	noun
893	potential	adj	consequence	noun	937	first	adj	meeting	noun
894	key	adj	difference	noun	938	brief	adj	overview	noun
895	different	adj	discipline	noun	939	social	adj	power	noun
896	clear	adj	distinction	noun	940	complex	adj	problem	noun
897	further	adj	exploration	noun	941	same	adj	question	noun
898	major	adj	factor	noun	942	common	adj	theme	noun
899	personal	adj	interest	noun	943	nuanced	adj	understanding	noun
900	deep	adj	level	noun	944	widespread	adj	use	noun
901	consistent	adj	pattern	noun	945	unique	adj	contribution	noun
902	broad	adj	perspective	noun	946	future	adj	development	noun
903	big	adj	picture	noun	947	different	adj	feature	noun
904	traditional	adj	role	noun	948	new	adj	framework	noun
905	social	adj	setting	noun	949	ultimate	adj	goal	noun
906	more	adj	study	noun	950	strong	adj	interest	noun
907	formal	adj	system	noun	951	different	adj	issue	noun
908	other	adj	theory	noun	952	own	adj	need	noun
909	same	adj	vein	noun	953	new	adj	source	noun
910	long	adj	way	noun	954	similar	adj	structure	noun
911	western	adj	world	noun	955	key	adj	theme	noun
912	individual	adj	behavior	noun	956	great	adj	variety	noun
913	substantial	adj	change	noun	957	own	adj	view	noun
914	scientific	adj	community	noun	958	specific	adj	action	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
959	cumulative	adj	effect	noun	1003	subtle	adj	difference	noun
960	important	adj	event	noun	1004	first	adj	dimension	noun
961	main	adj	objective	noun	1005	more	adj	emphasis	noun
962	essential	adj	part	noun	1006	only	adj	exception	noun
963	social	adj	perspective	noun	1007	great	adj	focus	noun
964	second	adj	reason	noun	1008	other	adj	influence	noun
965	final	adj	stage	noun	1009	horizontal	adj	line	noun
966	current	adj	theory	noun	1010	considerable	adj	number	noun
967	political	adj	view	noun	1011	large	adj	portion	noun
968	first	adj	category	noun	1012	new	adj	position	noun
969	new	adj	category	noun	1013	critical	adj	question	noun
970	primary	adj	concern	noun	1014	different	adj	school	noun
971	less	adj	degree	noun	1015	complex	adj	structure	noun
972	active	adj	engagement	noun	1016	local	adj	area	noun
973	few	adj	example	noun	1017	second	adj	category	noun
974	particular	adj	focus	noun	1018	significant	adj	challenge	noun
975	such	adj	group	noun	1019	same	adj	condition	noun
976	little	adj	information	noun	1020	current	adj	debate	noun
977	unique	adj	opportunity	noun	1021	psychological	adj	factor	noun
978	final	adj	set	noun	1022	particular	adj	identity	noun
979	necessary	adj	step	noun	1023	particular	adj	importance	noun
980	various	adj	study	noun	1024	long	adj	line	noun
981	particular	adj	time	noun	1025	primary	adj	objective	noun
982	many	adj	type	noun	1026	significant	adj	part	noun
983	early	adj	version	noun	1027	first	adj	point	noun
984	significant	adj	way	noun	1028	different	adj	purpose	noun
985	same	adj	degree	noun	1029	explicit	adj	reference	noun
986	essential	adj	element	noun	1030	possible	adj	response	noun
987	specific	adj	event	noun	1031	other	adj	service	noun
988	recent	adj	example	noun	1032	outside	adj	world	noun
989	different	adj	identity	noun	1033	certain	adj	amount	noun
990	central	adj	issue	noun	1034	other	adj	attribute	noun
991	small	adj	part	noun	1035	different	adj	community	noun
992	complete	adj	picture	noun	1036	great	adj	concern	noun
993	many	adj	reason	noun	1037	present	adj	context	noun
994	large	adj	role	noun	1038	narrow	adj	definition	noun
995	full	adj	scale	noun	1039	large	adj	degree	noun
996	broad	adj	spectrum	noun	1040	historical	adj	development	noun
997	general	adj	tendency	noun	1041	temporal	adj	dimension	noun
998	clear	adj	understanding	noun	1042	such	adj	example	noun
999	complex	adj	way	noun	1043	common	adj	form	noun
1000	great	adj	ability	noun	1044	distinct	adj	group	noun
1001	detailed	adj	account	noun	1045	other	adj	indicator	noun
1002	basic	adj	assumption	noun	1046	such	adj	instance	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1047	such	adj	intervention	noun	1091	productive	adj	use	noun
1048	new	adj	legislation	noun	1092	such	adj	value	noun
1049	specific	adj	meaning	noun	1093	great	adj	weight	noun
1050	similar	adj	problem	noun	1094	direct	adj	access	noun
1051	theoretical	adj	understanding	noun	1095	possible	adj	answer	noun
1052	positive	adj	way	noun	1096	general	adj	approach	noun
1053	key	adj	word	noun	1097	major	adj	challenge	noun
1054	easy	adj	access	noun	1098	cognitive	adj	component	noun
1055	second	adj	approach	noun	1099	immediate	adj	context	noun
1056	different	adj	circumstance	noun	1100	recent	adj	discussion	noun
1057	various	adj	element	noun	1101	different	adj	element	noun
1058	other	adj	phenomenon	noun	1102	specific	adj	example	noun
1059	specific	adj	problem	noun	1103	distinct	adj	form	noun
1060	large	adj	quantity	noun	1104	particular	adj	individual	noun
1061	more	adj	question	noun	1105	same	adj	individual	noun
1062	subsequent	adj	section	noun	1106	different	adj	organization	noun
1063	new	adj	structure	noun	1107	particular	adj	point	noun
1064	powerful	adj	tool	noun	1108	second	adj	point	noun
1065	own	adj	understanding	noun	1109	such	adj	problem	noun
1066	symbolic	adj	value	noun	1110	primary	adj	purpose	noun
1067	substantial	adj	variation	noun	1111	similar	adj	situation	noun
1068	alternative	adj	view	noun	1112	alternative	adj	strategy	noun
1069	comprehensive	adj	analysis	noun	1113	various	adj	strategy	noun
1070	emotional	adj	attachment	noun	1114	strong	adj	tendency	noun
1071	horizontal	adj	axis	noun	1115	central	adj	theme	noun
1072	appropriate	adj	behavior	noun	1116	analytical	adj	tool	noun
1073	essential	adj	component	noun	1117	important	adj	tool	noun
1074	main	adj	concern	noun	1118	strategic	adj	use	noun
1075	early	adj	development	noun	1119	traditional	adj	view	noun
1076	such	adj	distinction	noun	1120	considerable	adj	amount	noun
1077	prominent	adj	example	noun	1121	key	adj	area	noun
1078	specific	adj	experience	noun	1122	careful	adj	attention	noun
1079	latter	adj	group	noun	1123	major	adj	component	noun
1080	major	adj	group	noun	1124	central	adj	element	noun
1081	great	adj	length	noun	1125	causal	adj	explanation	noun
1082	long	adj	list	noun	1126	little	adj	interest	noun
1083	recent	adj	past	noun	1127	common	adj	knowledge	noun
1084	different	adj	place	noun	1128	few	adj	minute	noun
1085	several	adj	point	noun	1129	standard	adj	practice	noun
1086	different	adj	relationship	noun	1130	particular	adj	setting	noun
1087	different	adj	response	noun	1131	possible	adj	solution	noun
1088	different	adj	size	noun	1132	comprehensive	adj	understanding	noun
1089	new	adj	study	noun	1133	small	adj	unit	noun
1090	cultural	adj	understanding	noun	1134	own	adj	value	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1135	recent	adj	article	noun	1179	theoretical	adj	concept	noun
1136	brief	adj	description	noun	1180	illustrative	adj	example	noun
1137	opposite	adj	end	noun	1181	empirical	adj	observation	noun
1138	common	adj	experience	noun	1182	general	adj	problem	noun
1139	central	adj	feature	noun	1183	own	adj	role	noun
1140	several	adj	group	noun	1184	other	adj	topic	noun
1141	broad	adj	implication	noun	1185	new	adj	world	noun
1142	same	adj	issue	noun	1186	careful	adj	analysis	noun
1143	several	adj	issue	noun	1187	early	adj	analysis	noun
1144	central	adj	part	noun	1188	substantial	adj	body	noun
1145	complex	adj	pattern	noun	1189	major	adj	concern	noun
1146	same	adj	point	noun	1190	close	adj	connection	noun
1147	specific	adj	point	noun	1191	crucial	adj	element	noun
1148	various	adj	practice	noun	1192	indirect	adj	evidence	noun
1149	low	adj	rating	noun	1193	substantial	adj	evidence	noun
1150	practical	adj	reason	noun	1194	central	adj	focus	noun
1151	general	adj	sense	noun	1195	exclusive	adj	focus	noun
1152	specific	adj	term	noun	1196	high	adj	importance	noun
1153	distinct	adj	type	noun	1197	controversial	adj	issue	noun
1154	great	adj	access	noun	1198	specific	adj	kind	noun
1155	particular	adj	concern	noun	1199	political	adj	landscape	noun
1156	core	adj	element	noun	1200	causal	adj	link	noun
1157	little	adj	experience	noun	1201	first	adj	period	noun
1158	essential	adj	feature	noun	1202	clear	adj	picture	noun
1159	other	adj	function	noun	1203	similar	adj	point	noun
1160	such	adj	individual	noun	1204	additional	adj	question	noun
1161	good	adj	intention	noun	1205	vital	adj	role	noun
1162	same	adj	method	noun	1206	easy	adj	task	noun
1163	certain	adj	number	noun	1207	difficult	adj	time	noun
1164	practical	adj	problem	noun	1208	collaborative	adj	work	noun
1165	main	adj	purpose	noun	1209	privileged	adj	access	noun
1166	further	adj	question	noun	1210	other	adj	account	noun
1167	little	adj	room	noun	1211	central	adj	component	noun
1168	such	adj	rule	noun	1212	other	adj	datum	noun
1169	specific	adj	set	noun	1213	many	adj	feature	noun
1170	first	adj	sight	noun	1214	new	adj	field	noun
1171	different	adj	solution	noun	1215	broad	adj	framework	noun
1172	same	adj	source	noun	1216	third	adj	group	noun
1173	various	adj	stage	noun	1217	major	adj	impact	noun
1174	overall	adj	structure	noun	1218	close	adj	inspection	noun
1175	detailed	adj	study	noun	1219	ongoing	adj	interaction	noun
1176	common	adj	way	noun	1220	other	adj	interest	noun
1177	strong	adj	case	noun	1221	effective	adj	mean	noun
1178	dramatic	adj	change	noun	1222	various	adj	model	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1223	alternative	adj	perspective	noun	1267	single	adj	individual	noun
1224	such	adj	phenomenon	noun	1268	large	adj	institution	noun
1225	theoretical	adj	point	noun	1269	daily	adj	interaction	noun
1226	historical	adj	root	noun	1270	various	adj	mechanism	noun
1227	opposite	adj	side	noun	1271	alternative	adj	method	noun
1228	similar	adj	strategy	noun	1272	social	adj	need	noun
1229	first	adj	task	noun	1273	contemporary	adj	practice	noun
1230	little	adj	value	noun	1274	compelling	adj	reason	noun
1231	practical	adj	application	noun	1275	additional	adj	source	noun
1232	geographical	adj	area	noun	1276	different	adj	story	noun
1233	recent	adj	call	noun	1277	main	adj	theme	noun
1234	new	adj	culture	noun	1278	common	adj	thread	noun
1235	prime	adj	example	noun	1279	central	adj	concept	noun
1236	first	adj	experience	noun	1280	similar	adj	concern	noun
1237	simple	adj	form	noun	1281	same	adj	conclusion	noun
1238	useful	adj	framework	noun	1282	traditional	adj	culture	noun
1239	local	adj	language	noun	1283	many	adj	dimension	noun
1240	available	adj	opportunity	noun	1284	academic	adj	discourse	noun
1241	various	adj	point	noun	1285	main	adj	group	noun
1242	real	adj	problem	noun	1286	central	adj	importance	noun
1243	good	adj	quality	noun	1287	high	adj	incidence	noun
1244	other	adj	respect	noun	1288	new	adj	initiative	noun
1245	direct	adj	result	noun	1289	specific	adj	need	noun
1246	specific	adj	setting	noun	1290	other	adj	position	noun
1247	several	adj	source	noun	1291	such	adj	relation	noun
1248	central	adj	tenet	noun	1292	own	adj	research	noun
1249	practical	adj	term	noun	1293	own	adj	resource	noun
1250	broad	adj	understanding	noun	1294	certain	adj	situation	noun
1251	large	adj	volume	noun	1295	same	adj	strategy	noun
1252	several	adj	week	noun	1296	other	adj	technique	noun
1253	other	adj	action	noun	1297	similar	adj	type	noun
1254	active	adj	agent	noun	1298	new	adj	version	noun
1255	first	adj	approach	noun	1299	general	adj	agreement	noun
1256	common	adj	assumption	noun	1300	dominant	adj	approach	noun
1257	strong	adj	commitment	noun	1301	great	adj	awareness	noun
1258	effective	adj	communication	noun	1302	wide	adj	community	noun
1259	such	adj	concept	noun	1303	different	adj	conclusion	noun
1260	further	adj	consideration	noun	1304	direct	adj	connection	noun
1261	theoretical	adj	consideration	noun	1305	careful	adj	consideration	noun
1262	recent	adj	debate	noun	1306	considerable	adj	degree	noun
1263	early	adj	discussion	noun	1307	crucial	adj	difference	noun
1264	little	adj	effort	noun	1308	certain	adj	element	noun
1265	particular	adj	event	noun	1309	particular	adj	emphasis	noun
1266	different	adj	facet	noun	1310	significant	adj	event	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1311	scientific	adj	evidence	noun	1355	specific	adj	individual	noun
1312	primary	adj	factor	noun	1356	viable	adj	option	noun
1313	other	adj	framework	noun	1357	particular	adj	outcome	noun
1314	clear	adj	indication	noun	1358	different	adj	picture	noun
1315	cultural	adj	influence	noun	1359	early	adj	point	noun
1316	personal	adj	interaction	noun	1360	similar	adj	position	noun
1317	different	adj	interest	noun	1361	minor	adj	role	noun
1318	social	adj	nature	noun	1362	complex	adj	situation	noun
1319	urgent	adj	need	noun	1363	common	adj	source	noun
1320	important	adj	one	noun	1364	negative	adj	stereotype	noun
1321	historical	adj	perspective	noun	1365	additional	adj	time	noun
1322	particular	adj	position	noun	1366	particular	adj	value	noun
1323	whole	adj	process	noun	1367	subsequent	adj	work	noun
1324	basic	adj	question	noun	1368	central	adj	area	noun
1325	new	adj	reality	noun	1369	such	adj	argument	noun
1326	interpersonal	adj	relation	noun	1370	structural	adj	aspect	noun
1327	great	adj	responsibility	noun	1371	general	adj	consensus	noun
1328	dramatic	adj	shift	noun	1372	significant	adj	degree	noun
1329	initial	adj	step	noun	1373	rapid	adj	development	noun
1330	conceptual	adj	tool	noun	1374	profound	adj	effect	noun
1331	full	adj	understanding	noun	1375	second	adj	line	noun
1332	wide	adj	variation	noun	1376	possible	adj	meaning	noun
1333	crucial	adj	aspect	noun	1377	ample	adj	opportunity	noun
1334	great	adj	change	noun	1378	current	adj	position	noun
1335	own	adj	choice	noun	1379	direct	adj	response	noun
1336	several	adj	feature	noun	1380	various	adj	role	noun
1337	various	adj	method	noun	1381	difficult	adj	situation	noun
1338	first	adj	page	noun	1382	new	adj	wave	noun
1339	other	adj	perspective	noun	1383	such	adj	consideration	noun
1340	complex	adj	picture	noun	1384	social	adj	construct	noun
1341	small	adj	portion	noun	1385	different	adj	family	noun
1342	personal	adj	preference	noun	1386	only	adj	form	noun
1343	well	adj	quality	noun	1387	same	adj	frame	noun
1344	different	adj	question	noun	1388	basic	adj	function	noun
1345	important	adj	task	noun	1389	primary	adj	function	noun
1346	second	adj	term	noun	1390	different	adj	idea	noun
1347	similar	adj	view	noun	1391	academic	adj	interest	noun
1348	wide	adj	audience	noun	1392	relevant	adj	issue	noun
1349	particular	adj	characteristic	noun	1393	well	adj	job	noun
1350	conscious	adj	effort	noun	1394	many	adj	kind	noun
1351	certain	adj	event	noun	1395	own	adj	problem	noun
1352	typical	adj	example	noun	1396	crucial	adj	question	noun
1353	main	adj	feature	noun	1397	more	adj	sense	noun
1354	other	adj	figure	noun	1398	many	adj	similarity	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1399	only	adj	source	noun	1443	dual	adj	role	noun
1400	same	adj	argument	noun	1444	different	adj	space	noun
1401	various	adj	combination	noun	1445	different	adj	status	noun
1402	less	adj	concern	noun	1446	other	adj	structure	noun
1403	social	adj	concern	noun	1447	multiple	adj	system	noun
1404	full	adj	detail	noun	1448	good	adj	time	noun
1405	pure	adj	form	noun	1449	little	adj	use	noun
1406	overall	adj	impression	noun	1450	fundamental	adj	way	noun
1407	several	adj	member	noun	1451	more	adj	way	noun
1408	last	adj	part	noun	1452	own	adj	analysis	noun
1409	third	adj	part	noun	1453	novel	adj	approach	noun
1410	central	adj	point	noun	1454	such	adj	assumption	noun
1411	difficult	adj	problem	noun	1455	most	adj	attention	noun
1412	fundamental	adj	problem	noun	1456	empirical	adj	basis	noun
1413	limited	adj	range	noun	1457	latter	adj	category	noun
1414	stable	adj	system	noun	1458	particular	adj	challenge	noun
1415	fundamental	adj	assumption	noun	1459	related	adj	concept	noun
1416	further	adj	attention	noun	1460	important	adj	context	noun
1417	major	adj	contribution	noun	1461	additional	adj	element	noun
1418	inherent	adj	difficulty	noun	1462	much	adj	emphasis	noun
1419	several	adj	element	noun	1463	extreme	adj	example	noun
1420	separate	adj	entity	noun	1464	solid	adj	foundation	noun
1421	simple	adj	fact	noun	1465	major	adj	implication	noun
1422	original	adj	form	noun	1466	new	adj	level	noun
1423	academic	adj	institution	noun	1467	different	adj	member	noun
1424	conceptual	adj	level	noun	1468	major	adj	reason	noun
1425	theoretical	adj	question	noun	1469	particular	adj	role	noun
1426	own	adj	set	noun	1470	social	adj	rule	noun
1427	new	adj	setting	noun	1471	such	adj	skill	noun
1428	other	adj	tool	noun	1472	third	adj	stage	noun
1429	own	adj	version	noun	1473	brief	adj	summary	noun
1430	direct	adj	way	noun	1474	several	adj	article	noun
1431	real	adj	change	noun	1475	same	adj	background	noun
1432	definitive	adj	conclusion	noun	1476	few	adj	change	noun
1433	new	adj	element	noun	1477	significant	adj	component	noun
1434	different	adj	line	noun	1478	short	adj	history	noun
1435	close	adj	link	noun	1479	diverse	adj	interest	noun
1436	different	adj	mean	noun	1480	early	adj	model	noun
1437	regular	adj	meeting	noun	1481	human	adj	need	noun
1438	relative	adj	merit	noun	1482	upper	adj	part	noun
1439	actual	adj	outcome	noun	1483	global	adj	perspective	noun
1440	other	adj	project	noun	1484	similar	adj	phenomenon	noun
1441	material	adj	resource	noun	1485	decisive	adj	role	noun
1442	more	adj	responsibility	noun	1486	practical	adj	way	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1487	same	adj	characteristic	noun	1531	also	adv	note	verb
1488	several	adj	form	noun	1532	widely	adv	used	verb
1489	explanatory	adj	framework	noun	1533	however	adv	remain	verb
1490	multiple	adj	interpretation	noun	1534	also	adv	report	verb
1491	detailed	adj	knowledge	noun	1535	commonly	adv	used	verb
1492	general	adj	lack	noun	1536	together	adv	work	verb
1493	slow	adj	pace	noun	1537	thus	adv	provide	verb
1494	such	adj	setting	noun	1538	seriously	adv	take	verb
1495	concrete	adj	term	noun	1539	together	adv	bring	verb
1496	indirect	adj	way	noun	1540	frequently	adv	used	verb
1497	myriad	adj	way	noun	1541	also	adv	examine	verb
1498	so	adv	do	verb	1542	then	adv	use	verb
1499	also	adv	see	verb	1543	also	adv	take	verb
1500	however	adv	be	verb	1544	back	adv	go	verb
1501	also	adv	find	verb	1545	below	adv	discuss	verb
1502	well	adv	know	verb	1546	well	adv	do	verb
1503	also	adv	include	verb	1547	however	adv	note	verb
1504	so	adv	call	verb	1548	however	adv	seem	verb
1505	also	adv	show	verb	1549	also	adv	reflect	verb
1506	well	adv	understand	verb	1550	also	adv	influence	verb
1507	also	adv	use	verb	1551	below	adv	see	verb
1508	also	adv	suggest	verb	1552	now	adv	turn	verb
1509	however	adv	have	verb	1553	however	adv	argue	verb
1510	also	adv	provide	verb	1554	also	adv	contribute	verb
1511	together	adv	take	verb	1555	however	adv	make	verb
1512	however	adv	find	verb	1556	also	adv	help	verb
1513	significantly	adv	differ	verb	1557	also	adv	affect	verb
1514	thus	adv	have	verb	1558	therefore	adv	have	verb
1515	however	adv	suggest	verb	1559	also	adv	indicate	verb
1516	often	adv	use	verb	1560	also	adv	reveal	verb
1517	well	adv	establish	verb	1561	also	adv	allow	verb
1518	well	adv	perform	verb	1562	well	adv	document	verb
1519	also	adv	make	verb	1563	here	adv	use	verb
1520	above	adv	discuss	verb	1564	also	adv	demonstrate	verb
1521	above	adv	mention	verb	1565	also	adv	need	verb
1522	well	adv	fit	verb	1566	significantly	adv	increase	verb
1523	however	adv	show	verb	1567	however	adv	provide	verb
1524	above	adv	describe	verb	1568	thus	adv	appear	verb
1525	far	adv	go	verb	1569	also	adv	serve	verb
1526	also	adv	consider	verb	1570	early	adv	noted	verb
1527	here	adv	present	verb	1571	thus	adv	seem	verb
1528	yet	adv	have	verb	1572	far	adv	explore	verb
1529	above	adv	noted	verb	1573	also	adv	observe	verb
1530	long	adv	stand	verb	1574	also	adv	appear	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1575	thus	adv	make	verb	1619	therefore	adv	use	verb
1576	also	adv	argue	verb	1620	early	adv	discuss	verb
1577	thus	adv	expect	verb	1621	immediately	adv	follow	verb
1578	also	adv	know	verb	1622	here	adv	discuss	verb
1579	however	adv	appear	verb	1623	also	adv	point	verb
1580	thus	adv	become	verb	1624	also	adv	ask	verb
1581	however	adv	reveal	verb	1625	well	adv	describe	verb
1582	also	adv	support	verb	1626	also	adv	expect	verb
1583	directly	adv	related	verb	1627	early	adv	mention	verb
1584	then	adv	ask	verb	1628	closely	adv	related	verb
1585	also	adv	associate	verb	1629	however	adv	use	verb
1586	however	adv	become	verb	1630	also	adv	offer	verb
1587	also	adv	increase	verb	1631	long	adv	last	verb
1588	also	adv	seem	verb	1632	however	adv	mean	verb
1589	thus	adv	suggest	verb	1633	forward	adv	move	verb
1590	also	adv	become	verb	1634	also	adv	refer	verb
1591	also	adv	explain	verb	1635	exclusively	adv	focus	verb
1592	however	adv	indicate	verb	1636	similarly	adv	find	verb
1593	far	adv	reach	verb	1637	indeed	adv	show	verb
1594	primarily	adv	focus	verb	1638	often	adv	take	verb
1595	also	adv	lead	verb	1639	here	adv	describe	verb
1596	forward	adv	put	verb	1640	also	adv	draw	verb
1597	well	adv	work	verb	1641	together	adv	live	verb
1598	first	adv	have	verb	1642	also	adv	occur	verb
1599	however	adv	take	verb	1643	also	adv	tend	verb
1600	also	adv	apply	verb	1644	thus	adv	use	verb
1601	first	adv	use	verb	1645	also	adv	calle	verb
1602	also	adv	highlight	verb	1646	above	adv	outline	verb
1603	heavily	adv	rely	verb	1647	back	adv	trace	verb
1604	also	adv	require	verb	1648	fully	adv	understand	verb
1605	also	adv	involve	verb	1649	however	adv	differ	verb
1606	also	adv	play	verb	1650	often	adv	associate	verb
1607	also	adv	mean	verb	1651	together	adv	come	verb
1608	far	adv	examine	verb	1652	well	adv	develop	verb
1609	also	adv	identify	verb	1653	also	adv	differ	verb
1610	first	adv	examine	verb	1654	first	adv	provide	verb
1611	far	adv	suggest	verb	1655	however	adv	require	verb
1612	negatively	adv	affect	verb	1656	widely	adv	accepte	verb
1613	strongly	adv	associate	verb	1657	finally	adv	find	verb
1614	thus	adv	see	verb	1658	however	adv	consider	verb
1615	thus	adv	support	verb	1659	specifically	adv	examine	verb
1616	well	adv	explain	verb	1660	closely	adv	link	verb
1617	directly	adv	affect	verb	1661	far	adv	develop	verb
1618	however	adv	see	verb	1662	more	adv	focus	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1663	however	adv	include	verb	1707	thus	adv	create	verb
1664	moreover	adv	find	verb	1708	often	adv	find	verb
1665	then	adv	present	verb	1709	therefore	adv	need	verb
1666	first	adv	consider	verb	1710	often	adv	involve	verb
1667	also	adv	create	verb	1711	above	adv	see	verb
1668	well	adv	defined	verb	1712	thus	adv	take	verb
1669	thus	adv	represent	verb	1713	only	adv	use	verb
1670	also	adv	describe	verb	1714	significantly	adv	affect	verb
1671	specifically	adv	focus	verb	1715	furthermore	adv	find	verb
1672	forward	adv	look	verb	1716	thus	adv	indicate	verb
1673	simply	adv	put	verb	1717	however	adv	lead	verb
1674	also	adv	recognize	verb	1718	also	adv	raise	verb
1675	also	adv	seek	verb	1719	however	adv	report	verb
1676	instead	adv	focus	verb	1720	first	adv	show	verb
1677	therefore	adv	include	verb	1721	therefore	adv	suggest	verb
1678	then	adv	become	verb	1722	finally	adv	have	verb
1679	socially	adv	construct	verb	1723	thus	adv	lead	verb
1680	highly	adv	educate	verb	1724	still	adv	remain	verb
1681	already	adv	exist	verb	1725	explicitly	adv	state	verb
1682	indeed	adv	find	verb	1726	also	adv	address	verb
1683	thus	adv	include	verb	1727	also	adv	hold	verb
1684	naturally	adv	occur	verb	1728	first	adv	present	verb
1685	thus	adv	allow	verb	1729	finally	adv	show	verb
1686	also	adv	contain	verb	1730	far	adv	support	verb
1687	moreover	adv	have	verb	1731	later	adv	discuss	verb
1688	thus	adv	increase	verb	1732	also	adv	emphasize	verb
1689	often	adv	refer	verb	1733	generally	adv	have	verb
1690	widely	adv	vary	verb	1734	thus	adv	argue	verb
1691	also	adv	present	verb	1735	well	adv	capture	verb
1692	differently	adv	put	verb	1736	however	adv	demonstrate	verb
1693	moreover	adv	show	verb	1737	well	adv	serve	verb
1694	closely	adv	associate	verb	1738	then	adv	turn	verb
1695	far	adv	show	verb	1739	actively	adv	engage	verb
1696	therefore	adv	provide	verb	1740	thus	adv	help	verb
1697	also	adv	explore	verb	1741	frequently	adv	occur	verb
1698	often	adv	see	verb	1742	thus	adv	require	verb
1699	long	adv	taken	verb	1743	significantly	adv	vary	verb
1700	highly	adv	value	verb	1744	however	adv	observe	verb
1701	directly	adv	address	verb	1745	considerably	adv	vary	verb
1702	also	adv	depend	verb	1746	then	adv	examine	verb
1703	however	adv	know	verb	1747	thus	adv	find	verb
1704	previously	adv	mention	verb	1748	mainly	adv	focus	verb
1705	here	adv	consider	verb	1749	also	adv	imply	verb
1706	therefore	adv	consider	verb	1750	strongly	adv	influence	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1751	close	adv	come	verb	1795	thus	adv	show	verb
1752	also	adv	emerge	verb	1796	first	adv	appear	verb
1753	clearly	adv	show	verb	1797	also	adv	bring	verb
1754	however	adv	tend	verb	1798	directly	adv	influence	verb
1755	now	adv	consider	verb	1799	then	adv	provide	verb
1756	often	adv	make	verb	1800	thus	adv	remain	verb
1757	also	adv	result	verb	1801	also	adv	assume	verb
1758	apart	adv	set	verb	1802	significantly	adv	contribute	verb
1759	indeed	adv	suggest	verb	1803	far	adv	argue	verb
1760	moreover	adv	suggest	verb	1804	also	adv	facilitate	verb
1761	early	adv	describe	verb	1805	typically	adv	have	verb
1762	well	adv	educate	verb	1806	also	adv	link	verb
1763	here	adv	focus	verb	1807	directly	adv	link	verb
1764	away	adv	shift	verb	1808	thus	adv	serve	verb
1765	also	adv	add	verb	1809	only	adv	consider	verb
1766	thus	adv	reduce	verb	1810	finally	adv	use	verb
1767	also	adv	relate	verb	1811	first	adv	describe	verb
1768	thus	adv	contribute	verb	1812	however	adv	fail	verb
1769	also	adv	express	verb	1813	still	adv	need	verb
1770	furthermore	adv	have	verb	1814	also	adv	represent	verb
1771	also	adv	produce	verb	1815	closely	adv	resemble	verb
1772	well	adv	represent	verb	1816	instead	adv	use	verb
1773	far	adv	take	verb	1817	here	adv	argue	verb
1774	often	adv	describe	verb	1818	often	adv	assume	verb
1775	previously	adv	describe	verb	1819	however	adv	come	verb
1776	then	adv	show	verb	1820	short	adv	live	verb
1777	then	adv	compare	verb	1821	generally	adv	speak	verb
1778	previously	adv	discuss	verb	1822	first	adv	suggest	verb
1779	also	adv	encourage	verb	1823	also	adv	discuss	verb
1780	also	adv	extend	verb	1824	far	adv	discuss	verb
1781	significantly	adv	influence	verb	1825	then	adv	follow	verb
1782	often	adv	lead	verb	1826	hence	adv	have	verb
1783	furthermore	adv	suggest	verb	1827	first	adv	introduce	verb
1784	long	adv	have	verb	1828	previously	adv	note	verb
1785	wide	adv	range	verb	1829	thus	adv	reflect	verb
1786	also	adv	receive	verb	1830	significantly	adv	change	verb
1787	significantly	adv	reduce	verb	1831	generally	adv	consider	verb
1788	generally	adv	accept	verb	1832	next	adv	examine	verb
1789	thus	adv	consider	verb	1833	necessarily	adv	mean	verb
1790	clear	adv	cut	verb	1834	often	adv	consider	verb
1791	clearly	adv	define	verb	1835	then	adv	consider	verb
1792	here	adv	examine	verb	1836	only	adv	find	verb
1793	indeed	adv	have	verb	1837	also	adv	focus	verb
1794	counter	adv	run	verb	1838	especially	adv	give	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1839	therefore	adv	make	verb	1883	clearly	adv	indicate	verb
1840	again	adv	see	verb	1884	however	adv	limit	verb
1841	clearly	adv	see	verb	1885	even	adv	make	verb
1842	here	adv	show	verb	1886	more	adv	rely	verb
1843	similarly	adv	show	verb	1887	therefore	adv	seem	verb
1844	also	adv	understand	verb	1888	finally	adv	suggest	verb
1845	also	adv	acknowledge	verb	1889	also	adv	confirm	verb
1846	also	adv	benefit	verb	1890	also	adv	perform	verb
1847	also	adv	change	verb	1891	specifically	adv	suggest	verb
1848	back	adv	date	verb	1892	specifically	adv	design	verb
1849	also	adv	develop	verb	1893	also	adv	employ	verb
1850	also	adv	experience	verb	1894	again	adv	find	verb
1851	however	adv	give	verb	1895	thus	adv	focus	verb
1852	thus	adv	give	verb	1896	first	adv	include	verb
1853	similarly	adv	have	verb	1897	far	adv	increase	verb
1854	also	adv	illustrate	verb	1898	first	adv	make	verb
1855	often	adv	occur	verb	1899	indeed	adv	seem	verb
1856	then	adv	apply	verb	1900	even	adv	take	verb
1857	indeed	adv	argue	verb	1901	deeply	adv	embed	verb
1858	therefore	adv	conclude	verb	1902	however	adv	go	verb
1859	however	adv	continue	verb	1903	already	adv	mention	verb
1860	then	adv	discuss	verb	1904	also	adv	reduce	verb
1861	often	adv	focus	verb	1905	actively	adv	seek	verb
1862	therefore	adv	focus	verb	1906	fully	adv	explain	verb
1863	negatively	adv	impact	verb	1907	largely	adv	ignore	verb
1864	strongly	adv	suggest	verb	1908	often	adv	include	verb
1865	closely	adv	tie	verb	1909	here	adv	propose	verb
1866	specifically	adv	argue	verb	1910	often	adv	require	verb
1867	however	adv	change	verb	1911	often	adv	cite	verb
1868	also	adv	come	verb	1912	first	adv	need	verb
1869	however	adv	do	verb	1913	however	adv	offer	verb
1870	also	adv	exclude	verb	1914	also	adv	shape	verb
1871	then	adv	explain	verb	1915	also	adv	share	verb
1872	here	adv	find	verb	1916	specifically	adv	ask	verb
1873	together	adv	hold	verb	1917	then	adv	describe	verb
1874	extensively	adv	study	verb	1918	closely	adv	examine	verb
1875	also	adv	account	verb	1919	also	adv	introduce	verb
1876	however	adv	allow	verb	1920	however	adv	occur	verb
1877	however	adv	assume	verb	1921	thus	adv	offer	verb
1878	however	adv	believe	verb	1922	long	adv	recognize	verb
1879	also	adv	compare	verb	1923	again	adv	use	verb
1880	broadly	adv	define	verb	1924	closely	adv	align	verb
1881	long	adv	establish	verb	1925	therefore	adv	appear	verb
1882	then	adv	expect	verb	1926	similarly	adv	argue	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
1927	later	adv	become	verb	1971	well	adv	reflect	verb
1928	also	adv	enable	verb	1972	specifically	adv	relate	verb
1929	already	adv	establish	verb	1973	however	adv	represent	verb
1930	also	adv	exist	verb	1974	far	adv	reveal	verb
1931	first	adv	focus	verb	1975	however	adv	support	verb
1932	well	adv	have	verb	1976	sometimes	adv	use	verb
1933	however	adv	point	verb	1977	however	adv	acknowledge	verb
1934	high	adv	rank	verb	1978	nevertheless	adv	be	verb
1935	therefore	adv	represent	verb	1979	here	adv	define	verb
1936	effectively	adv	use	verb	1980	actively	adv	involve	verb
1937	far	adv	complicate	verb	1981	then	adv	move	verb
1938	thus	adv	constitute	verb	1982	here	adv	provide	verb
1939	however	adv	emerge	verb	1983	also	adv	rely	verb
1940	also	adv	inform	verb	1984	together	adv	suggest	verb
1941	however	adv	present	verb	1985	greatly	adv	vary	verb
1942	however	adv	produce	verb	1986	instead	adv	argue	verb
1943	therefore	adv	take	verb	1987	likely	adv	have	verb
1944	thus	adv	assume	verb	1988	necessarily	adv	have	verb
1945	briefly	adv	discuss	verb	1989	moreover	adv	make	verb
1946	largely	adv	focus	verb	1990	however	adv	raise	verb
1947	also	adv	incorporate	verb	1991	therefore	adv	require	verb
1948	only	adv	make	verb	1992	only	adv	understand	verb
1949	thus	adv	present	verb	1993	specifically	adv	address	verb
1950	far	adv	remove	verb	1994	then	adv	analyze	verb
1951	strictly	adv	speak	verb	1995	broad	adv	base	verb
1952	even	adv	use	verb	1996	therefore	adv	become	verb
1953	closely	adv	work	verb	1997	thus	adv	demonstrate	verb
1954	therefore	adv	assume	verb	1998	thus	adv	enable	verb
1955	substantially	adv	differ	verb	1999	therefore	adv	examine	verb
1956	far	adv	explain	verb	2000	also	adv	fail	verb
1957	far	adv	find	verb	2001	eventually	adv	lead	verb
1958	clearly	adv	have	verb	2002	then	adv	lead	verb
1959	however	adv	imply	verb	2003	thus	adv	limit	verb
1960	here	adv	make	verb	2004	accurately	adv	reflect	verb
1961	far	adv	reduce	verb	2005	however	adv	reflect	verb
1962	instead	adv	suggest	verb	2006	often	adv	result	verb
1963	successfully	adv	use	verb	2007	also	adv	begin	verb
1964	carefully	adv	consider	verb	2008	then	adv	begin	verb
1965	clearly	adv	demonstrate	verb	2009	also	adv	believe	verb
1966	partly	adv	explain	verb	2010	closely	adv	connect	verb
1967	together	adv	group	verb	2011	however	adv	emphasize	verb
1968	dramatically	adv	increase	verb	2012	fast	adv	grow	verb
1969	ultimately	adv	lead	verb	2013	yet	adv	know	verb
1970	above	adv	present	verb	2014	also	adv	limit	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2015	often	adv	overlook	verb	2059	increasingly	adv	become	verb
2016	sometimes	adv	refer	verb	2060	quickly	adv	become	verb
2017	well	adv	suit	verb	2061	often	adv	characterize	verb
2018	only	adv	become	verb	2062	then	adv	continue	verb
2019	just	adv	begin	verb	2063	fully	adv	develop	verb
2020	also	adv	challenge	verb	2064	short	adv	fall	verb
2021	less	adv	develop	verb	2065	also	adv	feel	verb
2022	also	adv	document	verb	2066	far	adv	have	verb
2023	also	adv	enhance	verb	2067	also	adv	improve	verb
2024	also	adv	establish	verb	2068	however	adv	lack	verb
2025	instead	adv	find	verb	2069	then	adv	leave	verb
2026	still	adv	find	verb	2070	indeed	adv	make	verb
2027	nevertheless	adv	have	verb	2071	still	adv	make	verb
2028	still	adv	hold	verb	2072	furthermore	adv	provide	verb
2029	closely	adv	look	verb	2073	moreover	adv	provide	verb
2030	closely	adv	match	verb	2074	however	adv	receive	verb
2031	then	adv	proceed	verb	2075	far	adv	strengthen	verb
2032	thus	adv	propose	verb	2076	first	adv	take	verb
2033	later	adv	return	verb	2077	far	adv	be	verb
2034	actually	adv	use	verb	2078	so	adv	be	verb
2035	above	adv	argue	verb	2079	indeed	adv	become	verb
2036	also	adv	carry	verb	2080	also	adv	characterize	verb
2037	also	adv	display	verb	2081	below	adv	explain	verb
2038	heavily	adv	draw	verb	2082	thus	adv	influence	verb
2039	back	adv	fall	verb	2083	often	adv	perceive	verb
2040	long	adv	hold	verb	2084	then	adv	perform	verb
2041	widely	adv	hold	verb	2085	also	adv	remain	verb
2042	then	adv	introduce	verb	2086	thus	adv	result	verb
2043	nevertheless	adv	remain	verb	2087	indeed	adv	see	verb
2044	also	adv	speak	verb	2088	clearly	adv	state	verb
2045	similarly	adv	suggest	verb	2089	also	adv	view	verb
2046	then	adv	add	verb	2090	equally	adv	apply	verb
2047	typically	adv	associate	verb	2091	close	adv	bring	verb
2048	together	adv	bind	verb	2092	dramatically	adv	change	verb
2049	even	adv	consider	verb	2093	later	adv	come	verb
2050	far	adv	consider	verb	2094	also	adv	exhibit	verb
2051	also	adv	engage	verb	2095	perhaps	adv	have	verb
2052	necessarily	adv	lead	verb	2096	simply	adv	have	verb
2053	inextricably	adv	link	verb	2097	well	adv	inform	verb
2054	back	adv	move	verb	2098	also	adv	obtain	verb
2055	commonly	adv	refer	verb	2099	often	adv	provide	verb
2056	often	adv	rely	verb	2100	however	adv	recognize	verb
2057	generally	adv	see	verb	2101	here	adv	refer	verb
2058	recently	adv	argue	verb	2102	first	adv	see	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2103	generally	adv	support	verb	2147	necessarily	adv	involve	verb
2104	interchangeably	adv	use	verb	2148	thereby	adv	make	verb
2105	still	adv	use	verb	2149	recently	adv	propose	verb
2106	well	adv	apply	verb	2150	simply	adv	reflect	verb
2107	also	adv	attempt	verb	2151	moreover	adv	reveal	verb
2108	hence	adv	be	verb	2152	then	adv	work	verb
2109	sometimes	adv	call	verb	2153	often	adv	base	verb
2110	ever	adv	change	verb	2154	thus	adv	begin	verb
2111	still	adv	exist	verb	2155	above	adv	cite	verb
2112	also	adv	follow	verb	2156	actually	adv	do	verb
2113	regardless	adv	have	verb	2157	however	adv	extend	verb
2114	however	adv	leave	verb	2158	even	adv	find	verb
2115	long	adv	live	verb	2159	well	adv	go	verb
2116	also	adv	propose	verb	2160	clearly	adv	identify	verb
2117	then	adv	return	verb	2161	still	adv	know	verb
2118	therefore	adv	see	verb	2162	far	adv	provide	verb
2119	thus	adv	seek	verb	2163	yet	adv	remain	verb
2120	again	adv	show	verb	2164	even	adv	seem	verb
2121	here	adv	take	verb	2165	also	adv	state	verb
2122	often	adv	think	verb	2166	thus	adv	tend	verb
2123	generally	adv	understand	verb	2167	rarely	adv	use	verb
2124	thus	adv	associate	verb	2168	thus	adv	view	verb
2125	similarly	adv	be	verb	2169	elsewhere	adv	argue	verb
2126	eventually	adv	become	verb	2170	differently	adv	behave	verb
2127	so	adv	become	verb	2171	typically	adv	consider	verb
2128	thus	adv	explain	verb	2172	considerably	adv	differ	verb
2129	prominently	adv	figure	verb	2173	first	adv	look	verb
2130	again	adv	have	verb	2174	therefore	adv	remain	verb
2131	however	adv	involve	verb	2175	first	adv	seem	verb
2132	well	adv	match	verb	2176	again	adv	suggest	verb
2133	thus	adv	play	verb	2177	generally	adv	agree	verb
2134	thus	adv	reveal	verb	2178	thus	adv	confirm	verb
2135	instead	adv	seek	verb	2179	first	adv	establish	verb
2136	moreover	adv	seem	verb	2180	far	adv	extend	verb
2137	directly	adv	speak	verb	2181	far	adv	illustrate	verb
2138	differently	adv	treat	verb	2182	also	adv	maintain	verb
2139	however	adv	base	verb	2183	finally	adv	make	verb
2140	well	adv	characterize	verb	2184	yet	adv	make	verb
2141	far	adv	demonstrate	verb	2185	much	adv	need	verb
2142	also	adv	determine	verb	2186	recently	adv	publish	verb
2143	thus	adv	develop	verb	2187	above	adv	quote	verb
2144	partially	adv	explain	verb	2188	therefore	adv	seek	verb
2145	yet	adv	find	verb	2189	yet	adv	suggest	verb
2146	prior	adv	have	verb	2190	commonly	adv	accept	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2191	then	adv	assign	verb	2235	even	adv	appear	verb
2192	even	adv	become	verb	2236	however	adv	arise	verb
2193	instead	adv	become	verb	2237	recently	adv	begin	verb
2194	only	adv	begin	verb	2238	however	adv	claim	verb
2195	also	adv	capture	verb	2239	thus	adv	come	verb
2196	however	adv	conclude	verb	2240	together	adv	consider	verb
2197	thus	adv	define	verb	2241	thereby	adv	create	verb
2198	heavily	adv	depend	verb	2242	however	adv	develop	verb
2199	thus	adv	describe	verb	2243	far	adv	elaborate	verb
2200	here	adv	employ	verb	2244	however	adv	express	verb
2201	thus	adv	enhance	verb	2245	commonly	adv	find	verb
2202	fully	adv	explore	verb	2246	generally	adv	focus	verb
2203	also	adv	generate	verb	2247	now	adv	go	verb
2204	often	adv	give	verb	2248	instead	adv	have	verb
2205	often	adv	ignore	verb	2249	most	adv	know	verb
2206	also	adv	impact	verb	2250	actually	adv	make	verb
2207	directly	adv	involve	verb	2251	generally	adv	regard	verb
2208	widely	adv	know	verb	2252	often	adv	represent	verb
2209	close	adv	move	verb	2253	deep	adv	root	verb
2210	early	adv	occur	verb	2254	repeatedly	adv	show	verb
2211	only	adv	show	verb	2255	already	adv	take	verb
2212	nevertheless	adv	suggest	verb	2256	instead	adv	take	verb
2213	generally	adv	take	verb	2257	thus	adv	add	verb
2214	poorly	adv	understand	verb	2258	also	adv	aim	verb
2215	thus	adv	understand	verb	2259	therefore	adv	allow	verb
2216	well	adv	use	verb	2260	largely	adv	base	verb
2217	moreover	adv	argue	verb	2261	also	adv	continue	verb
2218	commonly	adv	associate	verb	2262	just	adv	describe	verb
2219	certainly	adv	be	verb	2263	geographically	adv	disperse	verb
2220	gradually	adv	become	verb	2264	critically	adv	examine	verb
2221	directly	adv	connect	verb	2265	thus	adv	face	verb
2222	first	adv	define	verb	2266	however	adv	feel	verb
2223	also	adv	distinguish	verb	2267	however	adv	identify	verb
2224	thus	adv	emerge	verb	2268	specifically	adv	indicate	verb
2225	also	adv	entail	verb	2269	then	adv	need	verb
2226	thus	adv	hold	verb	2270	well	adv	recognize	verb
2227	however	adv	interpret	verb	2271	even	adv	require	verb
2228	thus	adv	involve	verb	2272	first	adv	seek	verb
2229	typically	adv	occur	verb	2273	widely	adv	acknowledge	verb
2230	above	adv	show	verb	2274	well	adv	address	verb
2231	far	adv	understand	verb	2275	indeed	adv	appear	verb
2232	furthermore	adv	use	verb	2276	still	adv	consider	verb
2233	also	adv	write	verb	2277	far	adv	contribute	verb
2234	far	adv	analyze	verb	2278	far	adv	describe	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2279	first	adv	enter	verb	2323	equally	adv	distribute	verb
2280	thus	adv	identify	verb	2324	therefore	adv	exclude	verb
2281	thus	adv	interpret	verb	2325	nevertheless	adv	find	verb
2282	strongly	adv	link	verb	2326	newly	adv	form	verb
2283	together	adv	link	verb	2327	usually	adv	involve	verb
2284	frequently	adv	mention	verb	2328	inevitably	adv	lead	verb
2285	far	adv	note	verb	2329	also	adv	leave	verb
2286	also	adv	place	verb	2330	often	adv	link	verb
2287	also	adv	regard	verb	2331	most	adv	matter	verb
2288	long	adv	remain	verb	2332	well	adv	organize	verb
2289	first	adv	assess	verb	2333	largely	adv	overlook	verb
2290	thus	adv	continue	verb	2334	however	adv	propose	verb
2291	furthermore	adv	demonstrate	verb	2335	still	adv	provide	verb
2292	even	adv	do	verb	2336	forward	adv	push	verb
2293	really	adv	have	verb	2337	greatly	adv	reduce	verb
2294	also	adv	interpret	verb	2338	specifically	adv	refer	verb
2295	also	adv	operate	verb	2339	so	adv	remain	verb
2296	also	adv	promote	verb	2340	only	adv	report	verb
2297	also	adv	reinforce	verb	2341	then	adv	require	verb
2298	instead	adv	rely	verb	2342	also	adv	respond	verb
2299	nevertheless	adv	show	verb	2343	indeed	adv	reveal	verb
2300	thus	adv	address	verb	2344	then	adv	serve	verb
2301	furthermore	adv	argue	verb	2345	fully	adv	support	verb
2302	usually	adv	associate	verb	2346	finally	adv	take	verb
2303	also	adv	bear	verb	2347	so	adv	take	verb
2304	instead	adv	choose	verb	2348	then	adv	translate	verb
2305	greatly	adv	differ	verb	2349	often	adv	understand	verb
2306	narrowly	adv	focus	verb	2350	later	adv	use	verb
2307	actually	adv	happen	verb	2351	also	adv	analyse	verb
2308	here	adv	identify	verb	2352	here	adv	apply	verb
2309	clearly	adv	illustrate	verb	2353	early	adv	begin	verb
2310	here	adv	include	verb	2354	often	adv	conflict	verb
2311	above	adv	list	verb	2355	intimately	adv	connect	verb
2312	first	adv	note	verb	2356	then	adv	create	verb
2313	also	adv	pay	verb	2357	first	adv	demonstrate	verb
2314	thus	adv	refer	verb	2358	first	adv	give	verb
2315	exclusively	adv	rely	verb	2359	thus	adv	go	verb
2316	also	adv	stress	verb	2360	instead	adv	provide	verb
2317	similarly	adv	use	verb	2361	also	adv	question	verb
2318	already	adv	be	verb	2362	however	adv	reach	verb
2319	recently	adv	become	verb	2363	often	adv	regard	verb
2320	especially	adv	consider	verb	2364	however	adv	rely	verb
2321	similarly	adv	describe	verb	2365	then	adv	seek	verb
2322	already	adv	discuss	verb	2366	often	adv	serve	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2367	thus	adv	treat	verb	2411	thus	adv	do	verb
2368	easily	adv	understand	verb	2412	thus	adv	draw	verb
2369	however	adv	add	verb	2413	often	adv	emphasize	verb
2370	adequately	adv	address	verb	2414	later	adv	explain	verb
2371	only	adv	address	verb	2415	then	adv	explore	verb
2372	directly	adv	challenge	verb	2416	less	adv	focus	verb
2373	widely	adv	cite	verb	2417	importantly	adv	have	verb
2374	along	adv	go	verb	2418	rarely	adv	have	verb
2375	typically	adv	include	verb	2419	first	adv	involve	verb
2376	well	adv	integrate	verb	2420	specifically	adv	look	verb
2377	furthermore	adv	seem	verb	2421	then	adv	place	verb
2378	generally	adv	use	verb	2422	often	adv	produce	verb
2379	mainly	adv	use	verb	2423	often	adv	reflect	verb
2380	primarily	adv	base	verb	2424	always	adv	take	verb
2381	here	adv	become	verb	2425	however	adv	achieve	verb
2382	therefore	adv	call	verb	2426	thus	adv	avoid	verb
2383	adequately	adv	capture	verb	2427	then	adv	decide	verb
2384	usually	adv	consider	verb	2428	however	adv	define	verb
2385	thus	adv	depend	verb	2429	clearly	adv	distinguish	verb
2386	all	adv	encompass	verb	2430	firmly	adv	establish	verb
2387	here	adv	explore	verb	2431	thus	adv	extend	verb
2388	however	adv	fall	verb	2432	therefore	adv	involve	verb
2389	elsewhere	adv	find	verb	2433	commonly	adv	known	verb
2390	rather	adv	focus	verb	2434	however	adv	link	verb
2391	thus	adv	generate	verb	2435	however	adv	lose	verb
2392	ever	adv	increase	verb	2436	even	adv	mean	verb
2393	originally	adv	intend	verb	2437	then	adv	observe	verb
2394	also	adv	meet	verb	2438	also	adv	possess	verb
2395	first	adv	observe	verb	2439	accurately	adv	represent	verb
2396	however	adv	play	verb	2440	still	adv	require	verb
2397	well	adv	play	verb	2441	often	adv	seek	verb
2398	thus	adv	point	verb	2442	however	adv	share	verb
2399	first	adv	propose	verb	2443	clearly	adv	suggest	verb
2400	also	adv	prove	verb	2444	again	adv	take	verb
2401	far	adv	seem	verb	2445	therefore	adv	view	verb
2402	then	adv	set	verb	2446	often	adv	attribute	verb
2403	perhaps	adv	suggest	verb	2447	nonetheless	adv	be	verb
2404	rather	adv	use	verb	2448	constantly	adv	change	verb
2405	so	adv	use	verb	2449	thus	adv	explore	verb
2406	frequently	adv	cite	verb	2450	then	adv	extend	verb
2407	moreover	adv	consider	verb	2451	previously	adv	hold	verb
2408	also	adv	design	verb	2452	often	adv	identify	verb
2409	rapidly	adv	develop	verb	2453	even	adv	include	verb
2410	widely	adv	differ	verb	2454	often	adv	invoke	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2455	closely	adv	monitor	verb	2499	far	adv	pursue	verb
2456	instead	adv	need	verb	2500	clearly	adv	reflect	verb
2457	together	adv	piece	verb	2501	positively	adv	respond	verb
2458	nevertheless	adv	provide	verb	2502	often	adv	speak	verb
2459	well	adv	receive	verb	2503	also	adv	suffer	verb
2460	especially	adv	regard	verb	2504	previously	adv	take	verb
2461	only	adv	relate	verb	2505	long	adv	work	verb
2462	here	adv	represent	verb	2506	then	adv	act	verb
2463	still	adv	show	verb	2507	however	adv	adopt	verb
2464	moreover	adv	take	verb	2508	often	adv	adopt	verb
2465	separately	adv	treat	verb	2509	generally	adv	apply	verb
2466	however	adv	work	verb	2510	thus	adv	challenge	verb
2467	first	adv	apply	verb	2511	often	adv	cite	verb
2468	early	adv	argue	verb	2512	especially	adv	compare	verb
2469	then	adv	change	verb	2513	explicitly	adv	draw	verb
2470	instead	adv	consider	verb	2514	widely	adv	employ	verb
2471	also	adv	define	verb	2515	far	adv	establish	verb
2472	then	adv	enter	verb	2516	yet	adv	establish	verb
2473	even	adv	exist	verb	2517	perhaps	adv	explain	verb
2474	far	adv	expand	verb	2518	also	adv	form	verb
2475	well	adv	find	verb	2519	therefore	adv	give	verb
2476	closely	adv	identify	verb	2520	initially	adv	have	verb
2477	here	adv	mention	verb	2521	however	adv	highlight	verb
2478	also	adv	prevent	verb	2522	successfully	adv	manage	verb
2479	indeed	adv	provide	verb	2523	often	adv	need	verb
2480	even	adv	show	verb	2524	again	adv	provide	verb
2481	also	adv	start	verb	2525	sometimes	adv	see	verb
2482	regardless	adv	use	verb	2526	however	adv	state	verb
2483	actually	adv	work	verb	2527	still	adv	struggle	verb
2484	often	adv	apply	verb	2528	nonetheless	adv	suggest	verb
2485	early	adv	cite	verb	2529	however	adv	undermine	verb
2486	specifically	adv	consider	verb	2530	however	adv	view	verb
2487	often	adv	depend	verb	2531	therefore	adv	want	verb
2488	however	adv	display	verb	2532	still	adv	allow	verb
2489	only	adv	emerge	verb	2533	therefore	adv	believe	verb
2490	now	adv	include	verb	2534	emotionally	adv	charge	verb
2491	also	adv	invite	verb	2535	often	adv	compare	verb
2492	so	adv	know	verb	2536	greatly	adv	contribute	verb
2493	thus	adv	link	verb	2537	first	adv	draw	verb
2494	far	adv	make	verb	2538	back	adv	give	verb
2495	moreover	adv	need	verb	2539	also	adv	implement	verb
2496	actually	adv	occur	verb	2540	ahead	adv	look	verb
2497	therefore	adv	present	verb	2541	then	adv	meet	verb
2498	then	adv	propose	verb	2542	below	adv	outline	verb

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2543	also	adv	posit	verb	2587	very	adv	similar	adj
2544	often	adv	remain	verb	2588	relatively	adv	small	adj
2545	indeed	adv	represent	verb	2589	very	adv	few	adj
2546	ultimately	adv	result	verb	2590	more	adv	negative	adj
2547	thus	adv	share	verb	2591	very	adv	high	adj
2548	differently	adv	think	verb	2592	more	adv	detailed	adj
2549	commonly	adv	understand	verb	2593	more	adv	common	adj
2550	universally	adv	accept	verb	2594	more	adv	nuanced	adj
2551	thus	adv	act	verb	2595	more	adv	specific	adj
2552	also	adv	attribute	verb	2596	quite	adv	different	adj
2553	specifically	adv	describe	verb	2597	well	adv	suited	adj
2554	otherwise	adv	do	verb	2598	more	adv	successful	adj
2555	again	adv	emphasize	verb	2599	more	adv	pronounced	adj
2556	finally	adv	explain	verb	2600	more	adv	efficient	adj
2557	also	adv	mediate	verb	2601	more	adv	frequent	adj
2558	usually	adv	refer	verb	2602	most	adv	significant	adj
2559	early	adv	start	verb	2603	more	adv	comprehensive	adj
2560	furthermore	adv	take	verb	2604	most	adv	effective	adj
2561	rather	adv	take	verb	2605	particularly	adv	relevant	adj
2562	then	adv	try	verb	2606	most	adv	appropriate	adj
2563	first	adv	turn	verb	2607	more	adv	sensitive	adj
2564	more	adv	likely	adj	2608	especially	adv	important	adj
2565	statistically	adv	significant	adj	2609	much	adv	high	adj
2566	less	adv	likely	adj	2610	relatively	adv	stable	adj
2567	most	adv	important	adj	2611	most	adv	prominent	adj
2568	more	adv	important	adj	2612	relatively	adv	little	adj
2569	more	adv	complex	adj	2613	more	adv	powerful	adj
2570	more	adv	difficult	adj	2614	much	adv	large	adj
2571	very	adv	different	adj	2615	well	adv	able	adj
2572	more	adv	effective	adj	2616	more	adv	diverse	adj
2573	more	adv	positive	adj	2617	significantly	adv	more	adj
2574	significantly	adv	high	adj	2618	most	adv	relevant	adj
2575	most	adv	likely	adj	2619	more	adv	traditional	adj
2576	more	adv	recent	adj	2620	more	adv	familiar	adj
2577	significantly	adv	different	adj	2621	closely	adv	related	adj
2578	most	adv	recent	adj	2622	more	adv	appropriate	adj
2579	relatively	adv	high	adj	2623	more	adv	active	adj
2580	more	adv	general	adj	2624	more	adv	relevant	adj
2581	most	adv	common	adj	2625	relatively	adv	large	adj
2582	too	adv	much	adj	2626	more	adv	complete	adj
2583	open	adv	ended	adj	2627	more	adv	robust	adj
2584	more	adv	accurate	adj	2628	very	adv	limited	adj
2585	only	adv	few	adj	2629	highly	adv	significant	adj
2586	particularly	adv	important	adj	2630	mutually	adv	exclusive	adj

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2631	more	adv	sophisticated	adj	2675	most	adv	basic	adj
2632	more	adv	flexible	adj	2676	just	adv	few	adj
2633	more	adv	advanced	adj	2677	increasingly	adv	important	adj
2634	more	adv	complicated	adj	2678	somewhat	adv	different	adj
2635	slightly	adv	different	adj	2679	more	adv	inclined	adj
2636	as	adv	important	adj	2680	more	adv	prominent	adj
2637	most	adv	influential	adj	2681	very	adv	specific	adj
2638	very	adv	least	adj	2682	more	adv	interested	adj
2639	less	adv	effective	adj	2683	more	adv	useful	adj
2640	equally	adv	important	adj	2684	particularly	adv	interesting	adj
2641	more	adv	open	adj	2685	less	adv	frequent	adj
2642	more	adv	attractive	adj	2686	more	adv	conservative	adj
2643	more	adv	experienced	adj	2687	more	adv	critical	adj
2644	less	adv	important	adj	2688	more	adv	consistent	adj
2645	more	adv	limited	adj	2689	more	adv	inclusive	adj
2646	relatively	adv	few	adj	2690	particularly	adv	strong	adj
2647	more	adv	willing	adj	2691	less	adv	able	adj
2648	as	adv	high	adj	2692	more	adv	distant	adj
2649	much	adv	small	adj	2693	relatively	adv	weak	adj
2650	much	adv	great	adj	2694	more	adv	concerned	adj
2651	more	adv	precise	adj	2695	more	adv	fundamental	adj
2652	very	adv	likely	adj	2696	more	adv	intense	adj
2653	less	adv	clear	adj	2697	especially	adv	TRUE	adj
2654	qualitatively	adv	different	adj	2698	highly	adv	relevant	adj
2655	most	adv	obvious	adj	2699	more	adv	rigorous	adj
2656	more	adv	significant	adj	2700	more	adv	stringent	adj
2657	more	adv	stable	adj	2701	as	adv	large	adj
2658	readily	adv	available	adj	2702	more	adv	moderate	adj
2659	more	adv	extensive	adj	2703	most	adv	striking	adj
2660	as	adv	likely	adj	2704	more	adv	comfortable	adj
2661	most	adv	powerful	adj	2705	less	adv	common	adj
2662	relatively	adv	short	adj	2706	more	adv	formal	adj
2663	particularly	adv	useful	adj	2707	more	adv	reliable	adj
2664	more	adv	vulnerable	adj	2708	enough	adv	strong	adj
2665	even	adv	great	adj	2709	much	adv	strong	adj
2666	slightly	adv	high	adj	2710	particularly	adv	TRUE	adj
2667	most	adv	salient	adj	2711	radically	adv	different	adj
2668	more	adv	subtle	adj	2712	more	adv	realistic	adj
2669	publicly	adv	available	adj	2713	very	adv	same	adj
2670	more	adv	severe	adj	2714	more	adv	influential	adj
2671	more	adv	direct	adj	2715	more	adv	interesting	adj
2672	more	adv	challenging	adj	2716	more	adv	meaningful	adj
2673	fundamentally	adv	different	adj	2717	even	adv	strong	adj
2674	more	adv	explicit	adj	2718	particularly	adv	difficult	adj

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2719	more	adv	focused	adj	2763	more	adv	dependent	adj
2720	particularly	adv	salient	adj	2764	increasingly	adv	difficult	adj
2721	relatively	adv	easy	adj	2765	more	adv	profound	adj
2722	far	adv	great	adj	2766	more	adv	typical	adj
2723	enough	adv	large	adj	2767	more	adv	widespread	adj
2724	more	adv	radical	adj	2768	very	adv	common	adj
2725	slightly	adv	more	adj	2769	primarily	adv	concerned	adj
2726	quite	adv	similar	adj	2770	second	adv	large	adj
2727	more	adv	abstract	adj	2771	more	adv	balanced	adj
2728	rather	adv	different	adj	2772	most	adv	efficient	adj
2729	most	adv	difficult	adj	2773	almost	adv	identical	adj
2730	completely	adv	different	adj	2774	more	adv	ambiguous	adj
2731	extremely	adv	difficult	adj	2775	more	adv	apparent	adj
2732	more	adv	prone	adj	2776	only	adv	available	adj
2733	more	adv	dynamic	adj	2777	virtually	adv	identical	adj
2734	nearly	adv	identical	adj	2778	less	adv	relevant	adj
2735	more	adv	valuable	adj	2779	more	adv	thorough	adj
2736	relatively	adv	new	adj	2780	as	adv	different	adj
2737	more	adv	problematic	adj	2781	highly	adv	influential	adj
2738	especially	adv	relevant	adj	2782	highly	adv	specific	adj
2739	most	adv	useful	adj	2783	as	adv	diverse	adj
2740	more	adv	serious	adj	2784	even	adv	less	adj
2741	more	adv	central	adj	2785	less	adv	obvious	adj
2742	particularly	adv	high	adj	2786	remarkably	adv	similar	adj
2743	more	adv	able	adj	2787	more	adv	susceptible	adj
2744	less	adv	willing	adj	2788	more	adv	engaged	adj
2745	very	adv	clear	adj	2789	only	adv	possible	adj
2746	most	adv	critical	adj	2790	most	adv	prevalent	adj
2747	markedly	adv	different	adj	2791	relatively	adv	simple	adj
2748	most	adv	notable	adj	2792	more	adv	compelling	adj
2749	as	adv	strong	adj	2793	increasingly	adv	complex	adj
2750	more	adv	substantial	adj	2794	less	adv	concerned	adj
2751	less	adv	successful	adj	2795	vastly	adv	different	adj
2752	more	adv	aware	adj	2796	sufficiently	adv	large	adj
2753	much	adv	broad	adj	2797	particularly	adv	suited	adj
2754	most	adv	extreme	adj	2798	somewhat	adv	surprising	adj
2755	most	adv	fundamental	adj	2799	most	adv	valuable	adj
2756	more	adv	immediate	adj	2800	largely	adv	absent	adj
2757	relatively	adv	limited	adj	2801	easily	adv	accessible	adj
2758	more	adv	visible	adj	2802	most	adv	consistent	adj
2759	more	adv	accessible	adj	2803	more	adv	evident	adj
2760	entirely	adv	new	adj	2804	especially	adv	strong	adj
2761	more	adv	elaborate	adj	2805	more	adv	suitable	adj
2762	more	adv	transparent	adj	2806	hardly	adv	surprising	adj

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2807	particularly	adv	interested	adj	2851	much	adv	close	adj
2808	equally	adv	likely	adj	2852	most	adv	crucial	adj
2809	even	adv	low	adj	2853	little	adv	more	adj
2810	relatively	adv	rare	adj	2854	more	adv	promising	adj
2811	more	adv	conventional	adj	2855	more	adv	varied	adj
2812	perhaps	adv	important	adj	2856	more	adv	concrete	adj
2813	relatively	adv	minor	adj	2857	comparatively	adv	little	adj
2814	very	adv	particular	adj	2858	more	adv	surprising	adj
2815	well	adv	known	adj	2859	have	verb	effect	noun
2816	more	adv	neutral	adj	2860	play	verb	role	noun
2817	more	adv	obvious	adj	2861	take	verb	place	noun
2818	most	adv	active	adj	2862	have	verb	impact	noun
2819	most	adv	direct	adj	2863	provide	verb	evidence	noun
2820	considerably	adv	high	adj	2864	have	verb	implication	noun
2821	more	adv	apt	adj	2865	make	verb	sense	noun
2822	very	adv	general	adj	2866	make	verb	decision	noun
2823	as	adv	great	adj	2867	provide	verb	support	noun
2824	considerably	adv	more	adj	2868	test	verb	hypothesis	noun
2825	just	adv	much	adj	2869	provide	verb	information	noun
2826	particularly	adv	problematic	adj	2870	pay	verb	attention	noun
2827	more	adv	lenient	adj	2871	answer	verb	question	noun
2828	most	adv	compelling	adj	2872	examine	verb	effect	noun
2829	more	adv	global	adj	2873	provide	verb	insight	noun
2830	so	adv	large	adj	2874	solve	verb	problem	noun
2831	quite	adv	likely	adj	2875	make	verb	use	noun
2832	too	adv	long	adj	2876	draw	verb	attention	noun
2833	more	adv	rapid	adj	2877	raise	verb	question	noun
2834	unusually	adv	high	adj	2878	find	verb	evidence	noun
2835	only	adv	little	adj	2879	shed	verb	light	noun
2836	too	adv	narrow	adj	2880	ask	verb	question	noun
2837	entirely	adv	clear	adj	2881	show	verb	effect	noun
2838	far	adv	complicated	adj	2882	address	verb	question	noun
2839	substantially	adv	different	adj	2883	use	verb	model	noun
2840	most	adv	immediate	adj	2884	have	verb	level	noun
2841	quite	adv	possible	adj	2885	use	verb	measure	noun
2842	more	adv	potent	adj	2886	give	verb	rise	noun
2843	especially	adv	useful	adj	2887	have	verb	potential	noun
2844	more	adv	constrained	adj	2888	have	verb	access	noun
2845	very	adv	rare	adj	2889	use	verb	method	noun
2846	very	adv	complex	adj	2890	address	verb	issue	noun
2847	largely	adv	due	adj	2891	provide	verb	opportunity	noun
2848	well	adv	equipped	adj	2892	have	verb	power	noun
2849	most	adv	interested	adj	2893	conduct	verb	analysis	noun
2850	as	adv	significant	adj	2894	have	verb	influence	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2895	have	verb	child	noun	2939	highlight	verb	importance	noun
2896	take	verb	form	noun	2940	use	verb	word	noun
2897	have	verb	relationship	noun	2941	use	verb	information	noun
2898	make	verb	contribution	noun	2942	ask	verb	participant	noun
2899	use	verb	strategy	noun	2943	have	verb	meaning	noun
2900	have	verb	consequence	noun	2944	take	verb	position	noun
2901	have	verb	experience	noun	2945	focus	verb	attention	noun
2902	use	verb	approach	noun	2946	fill	verb	gap	noun
2903	achieve	verb	goal	noun	2947	reveal	verb	difference	noun
2904	support	verb	hypothesis	noun	2948	see	verb	figure	noun
2905	find	verb	support	noun	2949	draw	verb	conclusion	noun
2906	examine	verb	relationship	noun	2950	provide	verb	basis	noun
2907	perform	verb	task	noun	2951	have	verb	time	noun
2908	spend	verb	time	noun	2952	address	verb	problem	noun
2909	take	verb	advantage	noun	2953	make	verb	claim	noun
2910	take	verb	part	noun	2954	offer	verb	opportunity	noun
2911	receive	verb	attention	noun	2955	have	verb	score	noun
2912	find	verb	difference	noun	2956	show	verb	pattern	noun
2913	use	verb	language	noun	2957	take	verb	role	noun
2914	show	verb	difference	noun	2958	have	verb	number	noun
2915	have	verb	opportunity	noun	2959	emphasize	verb	importance	noun
2916	tell	verb	story	noun	2960	improve	verb	performance	noun
2917	use	verb	term	noun	2961	provide	verb	explanation	noun
2918	make	verb	choice	noun	2962	exert	verb	influence	noun
2919	adopt	verb	approach	noun	2963	examine	verb	role	noun
2920	have	verb	value	noun	2964	explain	verb	difference	noun
2921	have	verb	difficulty	noun	2965	speak	verb	language	noun
2922	have	verb	interest	noun	2966	analyze	verb	datum	noun
2923	make	verb	difference	noun	2967	provide	verb	framework	noun
2924	receive	verb	support	noun	2968	conduct	verb	study	noun
2925	have	verb	advantage	noun	2969	report	verb	result	noun
2926	take	verb	action	noun	2970	mediate	verb	relationship	noun
2927	meet	verb	need	noun	2971	take	verb	risk	noun
2928	find	verb	way	noun	2972	produce	verb	effect	noun
2929	provide	verb	example	noun	2973	provide	verb	account	noun
2930	attend	verb	school	noun	2974	have	verb	limitation	noun
2931	meet	verb	criterion	noun	2975	have	verb	ability	noun
2932	have	verb	knowledge	noun	2976	use	verb	procedure	noun
2933	take	verb	step	noun	2977	make	verb	connection	noun
2934	take	verb	approach	noun	2978	develop	verb	understanding	noun
2935	offer	verb	insight	noun	2979	provide	verb	service	noun
2936	have	verb	history	noun	2980	examine	verb	impact	noun
2937	make	verb	judgment	noun	2981	use	verb	technique	noun
2938	have	verb	role	noun	2982	form	verb	part	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
2983	make	verb	effort	noun	3027	show	verb	relationship	noun
2984	occupy	verb	position	noun	3028	take	verb	view	noun
2985	have	verb	choice	noun	3029	use	verb	framework	noun
2986	provide	verb	datum	noun	3030	affect	verb	outcome	noun
2987	gain	verb	access	noun	3031	emphasize	verb	role	noun
2988	serve	verb	purpose	noun	3032	provide	verb	context	noun
2989	make	verb	distinction	noun	3033	play	verb	game	noun
2990	give	verb	opportunity	noun	3034	take	verb	responsibility	noun
2991	have	verb	chance	noun	3035	acquire	verb	knowledge	noun
2992	have	verb	resource	noun	3036	have	verb	responsibility	noun
2993	take	verb	account	noun	3037	increase	verb	number	noun
2994	create	verb	opportunity	noun	3038	make	verb	point	noun
2995	raise	verb	concern	noun	3039	attract	verb	attention	noun
2996	yield	verb	result	noun	3040	create	verb	environment	noun
2997	direct	verb	attention	noun	3041	address	verb	concern	noun
2998	reach	verb	level	noun	3042	contain	verb	information	noun
2999	give	verb	way	noun	3043	provide	verb	guidance	noun
3000	have	verb	control	noun	3044	support	verb	argument	noun
3001	pose	verb	challenge	noun	3045	create	verb	space	noun
3002	provide	verb	overview	noun	3046	explore	verb	effect	noun
3003	have	verb	association	noun	3047	face	verb	challenge	noun
3004	have	verb	problem	noun	3048	gather	verb	information	noun
3005	lend	verb	support	noun	3049	use	verb	tool	noun
3006	have	verb	capacity	noun	3050	maintain	verb	relationship	noun
3007	make	verb	case	noun	3051	identify	verb	factor	noun
3008	raise	verb	issue	noun	3052	provide	verb	way	noun
3009	make	verb	assumption	noun	3053	explore	verb	relationship	noun
3010	make	verb	prediction	noun	3054	highlight	verb	role	noun
3011	produce	verb	result	noun	3055	pose	verb	threat	noun
3012	call	verb	attention	noun	3056	serve	verb	function	noun
3013	have	verb	reason	noun	3057	support	verb	view	noun
3014	take	verb	course	noun	3058	include	verb	number	noun
3015	take	verb	perspective	noun	3059	offer	verb	explanation	noun
3016	conduct	verb	research	noun	3060	perform	verb	analysis	noun
3017	investigate	verb	relationship	noun	3061	discuss	verb	implication	noun
3018	make	verb	reference	noun	3062	follow	verb	pattern	noun
3019	develop	verb	skill	noun	3063	have	verb	connection	noun
3020	take	verb	value	noun	3064	examine	verb	extent	noun
3021	make	verb	change	noun	3065	have	verb	understanding	noun
3022	provide	verb	access	noun	3066	collect	verb	information	noun
3023	support	verb	idea	noun	3067	play	verb	part	noun
3024	enter	verb	market	noun	3068	use	verb	form	noun
3025	express	verb	concern	noun	3069	indicate	verb	difference	noun
3026	influence	verb	decision	noun	3070	understand	verb	relationship	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3071	develop	verb	theory	noun	3115	meet	verb	requirement	noun
3072	improve	verb	quality	noun	3116	gain	verb	understanding	noun
3073	show	verb	evidence	noun	3117	hold	verb	position	noun
3074	consider	verb	effect	noun	3118	provide	verb	resource	noun
3075	gain	verb	insight	noun	3119	stress	verb	importance	noun
3076	have	verb	authority	noun	3120	have	verb	benefit	noun
3077	have	verb	information	noun	3121	use	verb	number	noun
3078	pave	verb	way	noun	3122	discuss	verb	issue	noun
3079	reach	verb	conclusion	noun	3123	show	verb	level	noun
3080	use	verb	technology	noun	3124	use	verb	system	noun
3081	have	verb	structure	noun	3125	adopt	verb	perspective	noun
3082	examine	verb	change	noun	3126	create	verb	condition	noun
3083	have	verb	option	noun	3127	have	verb	goal	noun
3084	place	verb	emphasis	noun	3128	use	verb	theory	noun
3085	provide	verb	answer	noun	3129	reflect	verb	difference	noun
3086	receive	verb	treatment	noun	3130	adopt	verb	strategy	noun
3087	run	verb	risk	noun	3131	develop	verb	knowledge	noun
3088	achieve	verb	objective	noun	3132	give	verb	example	noun
3089	have	verb	relevance	noun	3133	find	verb	job	noun
3090	have	verb	root	noun	3134	have	verb	support	noun
3091	have	verb	tendency	noun	3135	make	verb	comparison	noun
3092	support	verb	claim	noun	3136	perform	verb	function	noun
3093	address	verb	gap	noun	3137	take	verb	shape	noun
3094	establish	verb	relationship	noun	3138	gain	verb	knowledge	noun
3095	leave	verb	room	noun	3139	provide	verb	point	noun
3096	make	verb	progress	noun	3140	require	verb	attention	noun
3097	pose	verb	problem	noun	3141	resolve	verb	conflict	noun
3098	present	verb	evidence	noun	3142	understand	verb	role	noun
3099	reach	verb	agreement	noun	3143	experience	verb	level	noun
3100	reduce	verb	number	noun	3144	have	verb	characteristic	noun
3101	share	verb	information	noun	3145	seek	verb	information	noun
3102	form	verb	basis	noun	3146	elicit	verb	response	noun
3103	indicate	verb	level	noun	3147	have	verb	skill	noun
3104	offer	verb	support	noun	3148	provide	verb	foundation	noun
3105	have	verb	function	noun	3149	give	verb	voice	noun
3106	pose	verb	question	noun	3150	offer	verb	evidence	noun
3107	produce	verb	outcome	noun	3151	understand	verb	process	noun
3108	conduct	verb	experiment	noun	3152	analyze	verb	effect	noun
3109	give	verb	meaning	noun	3153	demonstrate	verb	importance	noun
3110	have	verb	status	noun	3154	describe	verb	process	noun
3111	make	verb	argument	noun	3155	examine	verb	influence	noun
3112	provide	verb	mean	noun	3156	have	verb	intention	noun
3113	support	verb	notion	noun	3157	provide	verb	perspective	noun
3114	use	verb	criterion	noun	3158	have	verb	basis	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3159	meet	verb	demand	noun	3203	meet	verb	expectation	noun
3160	provide	verb	description	noun	3204	overcome	verb	problem	noun
3161	undergo	verb	change	noun	3205	increase	verb	level	noun
3162	achieve	verb	level	noun	3206	make	verb	statement	noun
3163	provide	verb	analysis	noun	3207	exercise	verb	control	noun
3164	assume	verb	role	noun	3208	present	verb	analysis	noun
3165	explore	verb	role	noun	3209	raise	verb	possibility	noun
3166	have	verb	dimension	noun	3210	resolve	verb	problem	noun
3167	provide	verb	justification	noun	3211	construct	verb	identity	noun
3168	use	verb	concept	noun	3212	have	verb	outcome	noun
3169	take	verb	effect	noun	3213	have	verb	property	noun
3170	explore	verb	possibility	noun	3214	increase	verb	chance	noun
3171	express	verb	view	noun	3215	meet	verb	standard	noun
3172	have	verb	need	noun	3216	propose	verb	model	noun
3173	have	verb	significance	noun	3217	show	verb	sign	noun
3174	limit	verb	ability	noun	3218	have	verb	size	noun
3175	make	verb	attempt	noun	3219	suggest	verb	way	noun
3176	shift	verb	focus	noun	3220	create	verb	problem	noun
3177	take	verb	look	noun	3221	follow	verb	rule	noun
3178	take	verb	measure	noun	3222	have	verb	evidence	noun
3179	have	verb	tie	noun	3223	raise	verb	awareness	noun
3180	identify	verb	pattern	noun	3224	show	verb	interest	noun
3181	leave	verb	question	noun	3225	address	verb	need	noun
3182	set	verb	stage	noun	3226	devote	verb	attention	noun
3183	understand	verb	nature	noun	3227	establish	verb	link	noun
3184	address	verb	challenge	noun	3228	examine	verb	way	noun
3185	create	verb	sense	noun	3229	have	verb	bearing	noun
3186	cross	verb	boundary	noun	3230	include	verb	term	noun
3187	follow	verb	approach	noun	3231	offer	verb	way	noun
3188	form	verb	relationship	noun	3232	open	verb	space	noun
3189	give	verb	attention	noun	3233	provide	verb	detail	noun
3190	have	verb	contact	noun	3234	reveal	verb	pattern	noun
3191	have	verb	feature	noun	3235	add	verb	value	noun
3192	influence	verb	behavior	noun	3236	afford	verb	opportunity	noun
3193	open	verb	possibility	noun	3237	have	verb	form	noun
3194	turn	verb	attention	noun	3238	have	verb	sense	noun
3195	bridge	verb	gap	noun	3239	receive	verb	training	noun
3196	give	verb	access	noun	3240	develop	verb	strategy	noun
3197	have	verb	view	noun	3241	draw	verb	inference	noun
3198	present	verb	challenge	noun	3242	identify	verb	type	noun
3199	find	verb	solution	noun	3243	receive	verb	information	noun
3200	give	verb	weight	noun	3244	require	verb	effort	noun
3201	have	verb	system	noun	3245	use	verb	variety	noun
3202	make	verb	mistake	noun	3246	affect	verb	behavior	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3247	attend	verb	meeting	noun	3291	develop	verb	approach	noun
3248	develop	verb	relationship	noun	3292	give	verb	account	noun
3249	examine	verb	question	noun	3293	give	verb	sense	noun
3250	highlight	verb	need	noun	3294	provide	verb	definition	noun
3251	introduce	verb	concept	noun	3295	provide	verb	indication	noun
3252	suggest	verb	need	noun	3296	provide	verb	protection	noun
3253	take	verb	stance	noun	3297	facilitate	verb	development	noun
3254	have	verb	background	noun	3298	have	verb	relation	noun
3255	provide	verb	model	noun	3299	influence	verb	development	noun
3256	reflect	verb	fact	noun	3300	take	verb	year	noun
3257	use	verb	example	noun	3301	assign	verb	value	noun
3258	use	verb	source	noun	3302	challenge	verb	assumption	noun
3259	cast	verb	doubt	noun	3303	cover	verb	range	noun
3260	have	verb	correlation	noun	3304	experience	verb	change	noun
3261	offer	verb	perspective	noun	3305	face	verb	problem	noun
3262	see	verb	discussion	noun	3306	involve	verb	process	noun
3263	strike	verb	balance	noun	3307	perform	verb	test	noun
3264	emphasize	verb	need	noun	3308	pursue	verb	interest	noun
3265	examine	verb	issue	noun	3309	share	verb	experience	noun
3266	explore	verb	impact	noun	3310	take	verb	control	noun
3267	gain	verb	support	noun	3311	underscore	verb	importance	noun
3268	have	verb	position	noun	3312	develop	verb	idea	noun
3269	influence	verb	choice	noun	3313	give	verb	impression	noun
3270	recognize	verb	importance	noun	3314	have	verb	feeling	noun
3271	send	verb	message	noun	3315	hold	verb	promise	noun
3272	confirm	verb	hypothesis	noun	3316	illustrate	verb	point	noun
3273	consider	verb	possibility	noun	3317	achieve	verb	success	noun
3274	explore	verb	question	noun	3318	apply	verb	approach	noun
3275	give	verb	information	noun	3319	have	verb	concern	noun
3276	have	verb	expectation	noun	3320	observe	verb	difference	noun
3277	reflect	verb	change	noun	3321	provide	verb	solution	noun
3278	satisfy	verb	need	noun	3322	provide	verb	space	noun
3279	send	verb	signal	noun	3323	reach	verb	goal	noun
3280	shift	verb	attention	noun	3324	take	verb	turn	noun
3281	accomplish	verb	goal	noun	3325	consider	verb	case	noun
3282	describe	verb	relationship	noun	3326	develop	verb	concept	noun
3283	draw	verb	distinction	noun	3327	enhance	verb	ability	noun
3284	have	verb	quality	noun	3328	face	verb	pressure	noun
3285	have	verb	way	noun	3329	follow	verb	logic	noun
3286	implement	verb	program	noun	3330	follow	verb	path	noun
3287	understand	verb	effect	noun	3331	form	verb	group	noun
3288	affect	verb	decision	noun	3332	offer	verb	account	noun
3289	bring	verb	change	noun	3333	possess	verb	power	noun
3290	create	verb	effect	noun	3334	rule	verb	possibility	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3335	change	verb	way	noun	3379	implement	verb	strategy	noun
3336	create	verb	context	noun	3380	take	verb	case	noun
3337	give	verb	support	noun	3381	experience	verb	difficulty	noun
3338	lay	verb	foundation	noun	3382	have	verb	purpose	noun
3339	present	verb	opportunity	noun	3383	keep	verb	track	noun
3340	put	verb	emphasis	noun	3384	make	verb	assessment	noun
3341	suggest	verb	possibility	noun	3385	offer	verb	advantage	noun
3342	take	verb	initiative	noun	3386	produce	verb	response	noun
3343	complete	verb	study	noun	3387	provide	verb	tool	noun
3344	make	verb	demand	noun	3388	take	verb	precedence	noun
3345	present	verb	picture	noun	3389	deepen	verb	understanding	noun
3346	show	verb	preference	noun	3390	demonstrate	verb	ability	noun
3347	add	verb	dimension	noun	3391	describe	verb	way	noun
3348	create	verb	value	noun	3392	develop	verb	set	noun
3349	exert	verb	control	noun	3393	emphasize	verb	nature	noun
3350	feel	verb	need	noun	3394	fill	verb	void	noun
3351	indicate	verb	degree	noun	3395	have	verb	belief	noun
3352	maintain	verb	position	noun	3396	reflect	verb	concern	noun
3353	take	verb	decision	noun	3397	satisfy	verb	requirement	noun
3354	adopt	verb	practice	noun	3398	set	verb	standard	noun
3355	explain	verb	change	noun	3399	understand	verb	difference	noun
3356	give	verb	time	noun	3400	understand	verb	impact	noun
3357	have	verb	trouble	noun	3401	use	verb	practice	noun
3358	identify	verb	difference	noun	3402	use	verb	study	noun
3359	reflect	verb	interest	noun	3403	begin	verb	process	noun
3360	take	verb	issue	noun	3404	facilitate	verb	process	noun
3361	achieve	verb	result	noun	3405	identify	verb	area	noun
3362	consider	verb	impact	noun	3406	maintain	verb	control	noun
3363	create	verb	situation	noun	3407	make	verb	contact	noun
3364	have	verb	friend	noun	3408	receive	verb	amount	noun
3365	have	verb	origin	noun	3409	require	verb	analysis	noun
3366	offer	verb	possibility	noun	3410	construct	verb	narrative	noun
3367	reveal	verb	relationship	noun	3411	contain	verb	element	noun
3368	show	verb	importance	noun	3412	effect	verb	change	noun
3369	deserve	verb	attention	noun	3413	explain	verb	process	noun
3370	employ	verb	strategy	noun	3414	explore	verb	issue	noun
3371	examine	verb	use	noun	3415	have	verb	record	noun
3372	have	verb	identity	noun	3416	need	verb	support	noun
3373	identify	verb	number	noun	3417	paint	verb	picture	noun
3374	identify	verb	strategy	noun	3418	pass	verb	test	noun
3375	present	verb	problem	noun	3419	place	verb	value	noun
3376	apply	verb	framework	noun	3420	present	verb	case	noun
3377	consider	verb	example	noun	3421	propose	verb	framework	noun
3378	have	verb	interaction	noun	3422	provide	verb	clue	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3423	pursue	verb	career	noun	3467	highlight	verb	difference	noun
3424	reach	verb	consensus	noun	3468	include	verb	range	noun
3425	reflect	verb	preference	noun	3469	increase	verb	understanding	noun
3426	share	verb	characteristic	noun	3470	influence	verb	way	noun
3427	support	verb	use	noun	3471	offer	verb	solution	noun
3428	consider	verb	issue	noun	3472	provide	verb	ground	noun
3429	give	verb	evidence	noun	3473	reach	verb	point	noun
3430	have	verb	strength	noun	3474	reflect	verb	experience	noun
3431	identify	verb	way	noun	3475	use	verb	mean	noun
3432	involve	verb	use	noun	3476	write	verb	letter	noun
3433	lack	verb	knowledge	noun	3477	consider	verb	implication	noun
3434	meet	verb	challenge	noun	3478	gain	verb	traction	noun
3435	produce	verb	form	noun	3479	provide	verb	illustration	noun
3436	provide	verb	education	noun	3480	use	verb	frame	noun
3437	provide	verb	summary	noun	3481	adopt	verb	view	noun
3438	reap	verb	benefit	noun	3482	explain	verb	phenomenon	noun
3439	require	verb	knowledge	noun	3483	face	verb	difficulty	noun
3440	share	verb	idea	noun	3484	garner	verb	support	noun
3441	spend	verb	money	noun	3485	highlight	verb	nature	noun
3442	understand	verb	way	noun	3486	illustrate	verb	way	noun
3443	use	verb	metaphor	noun	3487	indicate	verb	preference	noun
3444	use	verb	mode	noun	3488	suggest	verb	difference	noun
3445	broaden	verb	scope	noun	3489	demonstrate	verb	difference	noun
3446	cause	verb	problem	noun	3490	establish	verb	connection	noun
3447	convey	verb	message	noun	3491	illustrate	verb	effect	noun
3448	define	verb	identity	noun	3492	include	verb	element	noun
3449	employ	verb	method	noun	3493	provide	verb	sense	noun
3450	examine	verb	implication	noun	3494	represent	verb	type	noun
3451	have	verb	freedom	noun	3495	underline	verb	importance	noun
3452	indicate	verb	importance	noun	3496	carry	verb	implication	noun
3453	perform	verb	work	noun	3497	consider	verb	nature	noun
3454	present	verb	information	noun	3498	develop	verb	sense	noun
3455	show	verb	way	noun	3499	distinguish	verb	type	noun
3456	use	verb	instrument	noun	3500	encourage	verb	participation	noun
3457	examine	verb	practice	noun	3501	explain	verb	behavior	noun
3458	explore	verb	extent	noun	3502	follow	verb	lead	noun
3459	express	verb	desire	noun	3503	have	verb	desire	noun
3460	express	verb	interest	noun	3504	have	verb	limit	noun
3461	focus	verb	analysis	noun	3505	reach	verb	decision	noun
3462	give	verb	advice	noun	3506	reflect	verb	value	noun
3463	give	verb	chance	noun	3507	show	verb	number	noun
3464	give	verb	insight	noun	3508	study	verb	relationship	noun
3465	give	verb	name	noun	3509	take	verb	lead	noun
3466	have	verb	measure	noun	3510	use	verb	perspective	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3511	use	verb	tactic	noun	3555	develop	verb	perspective	noun
3512	allow	verb	comparison	noun	3556	have	verb	answer	noun
3513	emphasize	verb	aspect	noun	3557	have	verb	reputation	noun
3514	feel	verb	pressure	noun	3558	have	verb	source	noun
3515	give	verb	advantage	noun	3559	highlight	verb	aspect	noun
3516	have	verb	connotation	noun	3560	invest	verb	time	noun
3517	have	verb	point	noun	3561	provide	verb	link	noun
3518	occupy	verb	role	noun	3562	set	verb	limit	noun
3519	provide	verb	source	noun	3563	shape	verb	experience	noun
3520	achieve	verb	status	noun	3564	show	verb	degree	noun
3521	base	verb	decision	noun	3565	take	verb	context	noun
3522	carry	verb	weight	noun	3566	take	verb	opportunity	noun
3523	consider	verb	question	noun	3567	address	verb	aspect	noun
3524	encourage	verb	use	noun	3568	describe	verb	effect	noun
3525	extend	verb	analysis	noun	3569	establish	verb	rule	noun
3526	gain	verb	control	noun	3570	gain	verb	attention	noun
3527	have	verb	element	noun	3571	have	verb	motive	noun
3528	highlight	verb	fact	noun	3572	have	verb	result	noun
3529	highlight	verb	way	noun	3573	identify	verb	group	noun
3530	mark	verb	point	noun	3574	limit	verb	access	noun
3531	provide	verb	interpretation	noun	3575	promote	verb	change	noun
3532	recognize	verb	role	noun	3576	receive	verb	degree	noun
3533	see	verb	difference	noun	3577	see	verb	change	noun
3534	seek	verb	support	noun	3578	see	verb	evidence	noun
3535	yield	verb	insight	noun	3579	see	verb	value	noun
3536	address	verb	role	noun	3580	see	verb	world	noun
3537	affect	verb	way	noun	3581	tackle	verb	problem	noun
3538	construct	verb	meaning	noun	3582	adopt	verb	model	noun
3539	employ	verb	technique	noun	3583	establish	verb	relation	noun
3540	give	verb	indication	noun	3584	feel	verb	sense	noun
3541	recognize	verb	need	noun	3585	have	verb	approach	noun
3542	reflect	verb	level	noun	3586	have	verb	link	noun
3543	use	verb	terminology	noun	3587	include	verb	aspect	noun
3544	adopt	verb	position	noun	3588	inform	verb	understanding	noun
3545	challenge	verb	notion	noun	3589	question	verb	assumption	noun
3546	create	verb	structure	noun	3590	reflect	verb	view	noun
3547	describe	verb	situation	noun	3591	tackle	verb	issue	noun
3548	explore	verb	nature	noun	3592	avoid	verb	conflict	noun
3549	reduce	verb	difference	noun	3593	define	verb	boundary	noun
3550	reject	verb	notion	noun	3594	develop	verb	account	noun
3551	undergo	verb	process	noun	3595	enhance	verb	quality	noun
3552	achieve	verb	aim	noun	3596	expand	verb	scope	noun
3553	consider	verb	relationship	noun	3597	follow	verb	trajectory	noun
3554	create	verb	tension	noun	3598	generate	verb	response	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3599	have	verb	expertise	noun	3643	bring	verb	issue	noun
3600	have	verb	tool	noun	3644	create	verb	need	noun
3601	include	verb	form	noun	3645	discuss	verb	relationship	noun
3602	incorporate	verb	element	noun	3646	explore	verb	aspect	noun
3603	provide	verb	structure	noun	3647	inform	verb	policy	noun
3604	serve	verb	need	noun	3648	join	verb	group	noun
3605	voice	verb	concern	noun	3649	make	verb	plan	noun
3606	combine	verb	element	noun	3650	present	verb	argument	noun
3607	create	verb	category	noun	3651	propose	verb	approach	noun
3608	emphasize	verb	difference	noun	3652	reduce	verb	amount	noun
3609	include	verb	change	noun	3653	share	verb	value	noun
3610	involve	verb	degree	noun	3654	attain	verb	level	noun
3611	present	verb	example	noun	3655	attain	verb	status	noun
3612	present	verb	view	noun	3656	cast	verb	light	noun
3613	provide	verb	alternative	noun	3657	gain	verb	recognition	noun
3614	provide	verb	review	noun	3658	have	verb	room	noun
3615	reflect	verb	reality	noun	3659	illustrate	verb	relationship	noun
3616	reinforce	verb	idea	noun	3660	offer	verb	number	noun
3617	represent	verb	point	noun	3661	provide	verb	platform	noun
3618	address	verb	effect	noun	3662	reflect	verb	shift	noun
3619	create	verb	impression	noun	3663	apply	verb	technique	noun
3620	discuss	verb	problem	noun	3664	begin	verb	analysis	noun
3621	identify	verb	role	noun	3665	compare	verb	response	noun
3622	include	verb	reference	noun	3666	give	verb	option	noun
3623	present	verb	overview	noun	3667	have	verb	importance	noun
3624	provide	verb	argument	noun	3668	indicate	verb	need	noun
3625	reflect	verb	influence	noun	3669	introduce	verb	notion	noun
3626	reflect	verb	nature	noun	3670	involve	verb	group	noun
3627	reject	verb	idea	noun	3671	keep	verb	eye	noun
3628	require	verb	investment	noun	3672	receive	verb	approval	noun
3629	share	verb	feature	noun	3673	represent	verb	model	noun
3630	understand	verb	context	noun	3674	set	verb	parameter	noun
3631	use	verb	notion	noun	3675	shape	verb	process	noun
3632	consider	verb	difference	noun	3676	understand	verb	change	noun
3633	cover	verb	area	noun	3677	acknowledge	verb	limitation	noun
3634	describe	verb	role	noun	3678	allow	verb	time	noun
3635	encompass	verb	range	noun	3679	describe	verb	change	noun
3636	overcome	verb	obstacle	noun	3680	maintain	verb	focus	noun
3637	provide	verb	assessment	noun	3681	note	verb	importance	noun
3638	recognize	verb	value	noun	3682	offer	verb	advice	noun
3639	reveal	verb	nature	noun	3683	represent	verb	opportunity	noun
3640	serve	verb	goal	noun	3684	represent	verb	way	noun
3641	take	verb	challenge	noun	3685	take	verb	day	noun
3642	understand	verb	phenomenon	noun	3686	take	verb	task	noun

Sequence	Component I	POS I	Component II	POS II	Sequence	Component I	POS I	Component II	POS II
3687	adopt	verb	method	noun	3731	provide	verb	focus	noun
3688	follow	verb	tradition	noun	3732	reflect	verb	way	noun
3689	get	verb	sense	noun	3733	require	verb	examination	noun
3690	lose	verb	ability	noun	3734	suggest	verb	explanation	noun
3691	raise	verb	doubt	noun	3735	attract	verb	interest	noun
3692	take	verb	characteristic	noun	3736	base	verb	analysis	noun
3693	concern	verb	relationship	noun	3737	emphasize	verb	effect	noun
3694	consider	verb	range	noun	3738	foster	verb	relationship	noun
3695	constitute	verb	source	noun	3739	have	verb	process	noun
3696	follow	verb	practice	noun	3740	involve	verb	effort	noun
3697	have	verb	agency	noun	3741	offer	verb	assessment	noun
3698	have	verb	ramification	noun	3742	propose	verb	theory	noun
3699	identify	verb	model	noun	3743	reflect	verb	importance	noun
3700	include	verb	discussion	noun	3744	require	verb	explanation	noun
3701	reflect	verb	relationship	noun	3745	reveal	verb	tension	noun
3702	reinforce	verb	notion	noun	3746	consider	verb	change	noun
3703	require	verb	study	noun	3747	describe	verb	behavior	noun
3704	satisfy	verb	demand	noun	3748	see	verb	way	noun
3705	serve	verb	role	noun	3749	demonstrate	verb	need	noun
3706	understand	verb	pattern	noun	3750	describe	verb	state	noun
3707	blur	verb	distinction	noun	3751	find	verb	ground	noun
3708	enable	verb	people	noun	3752	merit	verb	attention	noun
3709	explain	verb	role	noun	3753	note	verb	difference	noun
3710	extend	verb	line	noun	3754	possess	verb	quality	noun
3711	face	verb	situation	noun	3755	present	verb	image	noun
3712	facilitate	verb	communication	noun	3756	understand	verb	cause	noun
3713	fulfil	verb	role	noun					
3714	have	verb	repercussion	noun					
3715	highlight	verb	effect	noun					
3716	initiate	verb	change	noun					
3717	involve	verb	people	noun					
3718	understand	verb	relation	noun					
3719	acknowledge	verb	role	noun					
3720	carry	verb	study	noun					
3721	consider	verb	aspect	noun					
3722	create	verb	network	noun					
3723	describe	verb	relation	noun					
3724	discuss	verb	aspect	noun					
3725	explore	verb	link	noun					
3726	give	verb	control	noun					
3727	illustrate	verb	difference	noun					
3728	include	verb	variety	noun					
3729	lack	verb	ability	noun					
3730	offer	verb	incentive	noun					

