

Corpus-Based Behavioral Profile Study of English Synonyms: Amazing, Astonishing, Stunning, Surprising, Startling, Staggering, and Shocking

Wenjing Wang*

Universiti Malaya, 50603 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia

* wangwenjing2052@gmail.com

Abstract: Employing the British National Corpus (BNC) as the source data and #LancsBox X as the tool, this study adopted a corpus-based behavioural profile approach to examine seven adjective near-synonyms (amazing, astonishing, stunning, surprising, startling, staggering, and shocking), focusing on the typical types of nouns they each modify and their register distribution patterns. The results showed that although these adjectives share a central meaning, they differ in usage patterns and formality levels, which complements the current inadequacies in describing these adjectives. Moreover, the study has confirmed the theory and applicability of the BP approach for investigating the semantic and usage patterns of synonyms in a set.

Keywords: Behavioral Profile, Near-Synonym, Distributional Pattern, Internal Semantic Structure, British National Corpus (BNC).

1. Introduction

Synonymy is an important, but also complex, language phenomenon (Divjak & Gries, 2006; Edmonds & Hirst, 2002; Taylor, 2002). Synonyms express the same or roughly the same meaning; however, they are not totally interchangeable because they don't have totally coinciding interpretations (Apresjan, 1995). According to DiMarco et al. (1993), while absolute synonymy (i.e. the substitutability of pairs of words in any situation) is rare, it is common for people to find sets of lexical items that are near synonyms in the sense that they have the same denotational meaning yet differ in respect of the connotational meanings and the collocational patterns. Such differences among near synonyms can sometimes cause problems for language learners. Therefore, it is important to know their semantic and usage nuances to help us convey our desired meanings to others effectively and efficiently.

Yet despite its great significance, so far, far less attention has been paid to near synonyms "in linguistics, psychology, lexicography, semantics, and computational linguistics than to the equally fundamental and much-studied polysemy" (Edmonds & Hirst, 2002). Since synonymy or near-synonymy is not widely understood among semantic relations (Taylor, 2002), more research needs to be carried out in order to illuminate this fascinating linguistic phenomenon, as well as to gain an in-depth understanding of how different specific groups of near synonyms function in terms of the internal semantic structure and usage patterns (Liu, 2010).

Against this background, the study aims to explore the semantic and stylistic differences of seven near-synonymous adjectives-amazing, astonishing, stunning, surprising, startling, staggering, and shocking-through a corpus-based Behavioural Profile (BP) approach. The reasons for investigating this set of near-synonyms by using the said approach are as follows: first, the internal semantic structure of this set of near-synonyms is rather complex; second, there is a lack of adequate descriptions of these synonyms involved; third, the effectiveness of corpus-based behavioral profile research on distinguishing synonymous adjectives needs to be

further tested.

The main research questions addressed in this study are:

- (1) What are the semantic variations among these near synonyms in terms of the typical types of nouns that they each modify?
- (2) What are the stylistic variations among these near synonyms in relation to their register distribution patterns?

2. Literature Review

2.1. The Definitions and Types of Synonyms

There are three types of synonyms, i.e., absolute synonyms, cognitive synonyms, and plesionyms or near synonyms. Absolute synonyms are a set of lexical items which can be substituted for one another in all linguistic contexts. It is widely acknowledged that absolute synonyms do not exist or at least are uncommon (Cruse, 1986; DiMarco et al., 1993; Divjak, 2010). Synonyms may be identical regarding central meanings but differ in peripheral meanings. Therefore, it is necessary to identify such differences among a set of synonyms to examine their internal semantic structure and usage patterns since "each of a set of synonyms has its particular value only because they stand in contrast with one another" (De Saussure, 1983, p. 114). Synonyms may vary in two different ways: "semantic and stylistic, or, equivalently, denotative and connotative" (DiMarco et al., 1993, p. 120). If the internal semantic structure of the set of synonyms is identical, replacing one with the other in a sentence will be viable because the denotation is the same. However, if the set of synonyms differs stylistically, substituting one for the other will change the connotation of the sentence.

Cruse (2000, pp. 158-161) made a distinction between "cognitive synonyms" and "plesionyms". Cognitive synonyms refer to lexical items which have the same propositional properties but the different expressive properties (Cruse, 1986, p. 273). Thus, one lexical item can be substituted for another in a sentence without changing the truth condition of the sentence. One of the examples is this set of verbs, i.e. "die", "pass away" and "kick the bucket" (Divjak,

2010). It is possible to state either “He passed away,” or “He kicked the bucket.” However, they cannot be used in one sentence, such as “He kicked the bucket but he did not die” (Divjak, 2010).

Plesionyms or near synonyms, which are more common in language than cognitive synonyms, refer to lexical items which differ not only in propositional properties but also in expressive properties. If one lexical item is replaced by another in a sentence, the truth condition of the sentence will be changed. One of the examples is this set of verbs, i.e., “kill” and “murder.” They can be used in one sentence, such as “He was killed, but I can assure you he was not murdered” (Divjak, 2010).

Considering that synonyms are very important in lexical semantics, it is a must for us to distinguish between sets of synonyms, especially under-researched synonymous adjectives, by identifying their internal semantic structure and distributional patterns.

2.2. The Complexity of the Internal Semantic Structure of the Near-synonym Set and the Inadequate Descriptions of the Near-synonym Set in the Existing Reference Materials

One of the main reasons for studying this set of near synonyms is its intricate internal semantic structure. For example, the seven synonymous adjectives have the same definitions in the dictionaries: “very surprising” and “difficult to believe” (Oxford Advanced American Dictionary, 2011; Longman Dictionary of Contemporary English, 2014; Oxford Advanced Learner’s Dictionary, 2015; Cambridge Advanced Learner’s Dictionary, 2015) and it seems that these adjectives can be interchangeable in modifying certain nouns such as thing, story and phenomenon. For example, we can say an amazing, astonishing, stunning, or a surprising, startling, staggering, and shocking fact or event as shown in the dictionaries. However, when modifying some other nouns, these adjectives cannot be substituted for one another. For instance, we can say “startling contrast” not “staggering contrast”, and “stunning victory” not “shocking victory” as demonstrated by the British National Corpus (BNC). It is not clear what the distributional patterns of these near synonyms are—that is, which types of nouns all seven adjectives can modify and which types of nouns only one or some of the adjectives can modify.

Another motivation for studying the seven synonymous adjectives is that there are inadequate descriptions of them in current reference materials. In addition to the four above-mentioned dictionaries, the other eleven reference materials I consulted are Longman Dictionary of American English (2014), The Oxford American College Dictionary (2002), The American Heritage Dictionary of the English Language (2015), Webster’s Essential Mini Dictionary (2011), The Newbury House Dictionary of American English (2000), The Cambridge Thesaurus of American English (1994), The American Heritage College Thesaurus (2004), The Concise Oxford American Thesaurus (2006), Webster’s New World Dictionary and Thesaurus (1996), Merriam-Webster’s Dictionary of Synonyms (1984), Longman Synonym Dictionary (1986). All these dictionaries offer general and even duplicated definitions for the seven synonyms, which may pose challenges for English language learners. For instance, “amazing” means very good, especially in

an unexpected way; “astonishing” means so surprising that it is difficult to believe; “astounding” means so surprising that it is almost impossible to believe; “surprising” means unusual or unexpected; “startling” means surprising and sometimes worrying; “staggering” means very shocking and surprising; “shocking” means very surprising, upsetting, and difficult to believe. Based on the definitions, we cannot ascertain the specific semantic and usage patterns of these adjectives from traditional reference materials. The inadequate descriptions of these adjectives are probably due to the major function and limited space of these reference materials. In fact, Liu (2010), in his study of six near-synonyms—chief, main, major, primary, and principal—also reported the inadequate descriptions and treatments of near-synonyms regarding the unique meanings and usage patterns of each of the adjectives. Furthermore, Church et al. (1994) found similar insufficient treatment of synonyms in their research of the synonyms ask for, demand, and request.

2.3. Behavioral Profile Approach Towards Corpus-based Lexical Semantics

The Behavioral Profile approach is a corpus-based approach towards lexical semantics (Gries, 2010), which assumes that the distributional patterns of a lexical item reveal many of its meanings and functions. It seems that Hanks (1996) is the first person to come up with the term “behavioral profile” to describe the distributional patterns of lexical items. The theoretical foundations underpinning this approach derived from the work by Firth and his followers Halliday and Sinclair (Liu, 2010).

According to Firth (1957, p. 11), “the complete meaning of a word is always contextual” and “you shall know a word by the company it keeps.” This famous assertion established the importance of collocation in the study of lexical semantics. For example, Halliday (1966, p. 156) argued that “it is the lexical restriction which is under focus: the extent to which an item is specified by its collocational environment.” Similarly, Sinclair (1991) concluded the description of word co-occurrence and advanced two different models of interpretation: the open-choice principle and the idiom principle. The open-choice principle, also called the “slot-and-filler” model, sees language as “a series of slots which have to be filled from a lexicon” (Sinclair, 1991). In other words, language users can select individual words to construct phrases

whose meanings are derived from the individual words (Liu, 2010). However, words do not occur randomly in a language, and they tend to collocate with each other. Thus, the idiom principle is built on the fact that language users can make use of “a large number of semi-constructed phrases that constitute single choices” (Sinclair, 1991).

Besides Firth, Halliday, and Sinclair, other scholars made theoretical contributions to the formation and development of the Behavioral Profile approach as well. For instance, Bolinger (1968, p. 127) noted that “a difference in syntactic form always spells a difference in meaning.” This is in line with the idea of Harris (1970, p. 785f.), who wrote “difference of meaning correlates with difference of distribution.” Moreover, Cruse (1986, p. 1) stated that “the semantic properties of a lexical item are fully reflected in appropriate aspects of the relations it contracts with actual and potential contexts.” These scholars worked together to establish a close link between lexical semantics and distributional patterns.

Guided by lexical-semantic theories as well as the BP

methodology, a lot of research on a group of (synonymous) verbs has emerged. Hanks (1996), in his corpus-based BP study of a set of verbs (urge, incite, bother, and abandon), investigated the complementation patterns (i.e. typical subjects, objects, and adverbials) of these verbs and the frequencies with which the various patterns occurred. The study found that the most common pattern of the verb “urge” is “a person urging another person to do something” (accounting for 61% of all occurrences in the Oxford Hector Corpus), while the least common use is “a person urging a horse or another person to move forward” (accounting for 3.5%). The findings are very informative and revealing because traditional dictionaries don’t give any clues about the distribution patterns of most polysemous words. Divjak (2006) used cluster analysis and the BP approach to study a set of synonymous verbs expressing “intend” in Russian. Using a sixteen-million-word literature section from the Amsterdam Corpus (Barentsen, 1999), the study systematically analyzed the similarities and differences among these verbs, created behavioral profiles for each of them, and identified the main semantic features that could differentiate them. Differentiating the semantic nuances of synonyms contributed to a clearer understanding of synonymy in linguistic research because sets of near-synonyms may share the same coarse-grained meaning but differ with respect to aspects of fine-grained meanings such as denotation, style, attitude, and usage (Edmonds & Hirst, 2002). Divjak & Gries (2006) examined 1585 instances of a set of near-synonymous Russian verbs that express “try” in the Amsterdam Corpus and the Russian National Corpus and revealed the internal structure of the near-synonym set in terms of their behavioral profiles, i. e., morphosyntactic, syntactic, and semantic features. This study verifies the assumed correlation between semantic and distributional properties of a set of near synonyms. Furthermore, Walker (2011) used the Bank of English (BoE) corpus, along with the commerce and finance data taken from British National Corpus, to establish a collocational profile for each of the three near synonymous verbs, i.e., RUN, MANAGE, and HEAD. The study showed that the typical objects of the verb “run” are non-human entities such as businesses, companies or firms, while the verb “manage” is more frequently associated with nouns referring to both non-human entities and people and the verb “head” tend to occur more frequently with nouns describing structures consisting of people, such as committees or panels. To sum up, the results of the above-mentioned studies on a set of (synonymous) verbs all show that the BP method is very effective in studying near synonyms (Divjak, 2006; Divjak & Gries, 2006; Gries, 2001; Gries & Otani, 2010; Hanks, 1996).

While most corpus-based behavioral studies focused on verbs, a few have focused on adjectives (Church et al, 1991; Gries, 2001; Gries & Otani, 2010). Church et al (1991, p. 119ff), for example, have shown that the synonymous adjectives strong and powerful differ greatly in terms of their significant collocates. The significance of these collocational patterns was evaluated by the mutual information measure, one of the oldest measures of collocation strength (Stefanowitsch, 2020). In other words, a word’s meaning can be defined and differentiated based on its (important) collocates. Gries (2001) presented a corpus-linguistic analysis of English -ic and -ical adjectives (e.g. logistic/logistical; symmetric/symmetrical). He applied the “Estimation of Significant Collocate Overlap” (ESCO) method to determine the degree of semantic similarity and differentiation between

the paired adjectives. The rationale behind this method is that on one hand, the more significant collocates word1 to word2 share, the more semantically similar they are, and on the other hand, the more discriminating collocates word1 to word2 have, the more semantically different they are. This implies that there is a close correlation between the meaning of lexical items and their collocational behavior (Kaunisto, 1999). Gries & Otani (2010) used the behavioral profile (BP) approach to explore the distributional characteristics of three synonymous adjectives: big, great, and large, and their antonymous set: little, small, and tiny based on the data from the British Component of the International Corpus of English. They first employed an R script to retrieve all instances of the lemmas of the six adjectives from the corpus, and then the data were imported into a spreadsheet software and annotated for a variety of properties referred to as ID tags (Atkins, 1987) such as the morphological, syntactic, semantic, and collocational characteristics. Then the resulting table consisting of 2,073 rows and 27 columns was statistically analyzed using the interactive R script BP 1.0 (Gries 2008), and the behavioral profiles of each adjective were thus established. Finally, they used hierarchical agglomerative cluster analysis to group words based on the similarity of their co-occurrence vectors. The study also validates the use of Behavioral Profiles (BP) as a corpus-based method for analyzing semantic similarity and differences among synonyms and antonyms.

The above review has suggested that the Behavioral Profiles (BP) approach lent itself well to the study of a set of near synonyms. Besides, the approach has primarily been used to examine sets of synonymous verbs, with only a few studies on sets of synonymous adjectives. As Hanks (1996, p. 92) explains, “there is no reason to believe that the procedures that help us to use and understand verbs are relevant to, say, nouns or adverbs.” Therefore, this study makes an important contribution by demonstrating the effectiveness of the BP approach in investigating a set of synonymous adjectives.

To summarize, the study aims to investigate the semantic and usage differences among the seven synonymous adjectives by focusing on their distributional patterns, especially the typical types of nouns that each of them modifies. In this study, I used the #LancsBox X tool due to its easy and free access, as well as its several built-in corpora (e.g. BNC 1994, BNC 2014, Brown Family) and user-friendly functions (e.g. KWIC, GraphColl, Words, and Texts). The corpus used for this study is the built-in 100 million-word British National Corpus (BNC) in #LancsBox X, containing texts from a variety of genres such as fiction, magazines, newspapers, academic texts, etc. The BNC was chosen because it is comprehensive, representative, and suitable for studying the lexical, grammatical, and semantic features of contemporary English.

3. Data Retrieval, Analysis, and Discussion

3.1. Overall Usage Patterns

To comprehend the semantic and usage patterns of the seven synonymous adjectives, we should probably start with their overall frequency patterns. I used the KWIC tool in #LancsBox X, which can generate a list of all instances of a search term in a corpus in the form of a concordance and find the frequency of a word or phrase in the corpus. The results are shown in table 1 below.

Table 1. Frequency of the seven adjectives

| | | | | | | |
|----------------|--------------|---------------|-----------------|--------------|-------------|----------------|
| stagger ing | shock ing | startli ng | astonis hing | stunn ing | amazi ng | surpris ing |
| 545 | 561 | 642 | 830 | 914 | 1,809 | 3,487 |

Table 1 displayed the frequency order of the seven adjectives from the lowest to the highest: astounding, staggering, shocking, startling, astonishing, amazing, and surprising, and huge differences in frequency among these adjectives (except for staggering and shocking, which only had a small difference in frequency).

However, the total frequency information alone does not give us any clues about the semantic and usage variations among these adjectives. Therefore, I consulted the fifteen dictionaries and thesauri as mentioned earlier, and it appears that the seven adjectives share a central meaning, i.e., “causing a feeling of great surprise or wonder”, although there are some variations in their source meanings, which may lead to some significant variations in their semantic and usage patterns, especially in the types of nouns each of the adjectives modifies.

To identify the types of nouns that are typically modified by each adjective, I used the GraphColl tool in #LancsBox X to search for the noun collocates of each adjective based on two measures: (1) raw frequency of co-occurrence with the adjectives; and (2) Mutual Information (MI). On one hand, the frequency measure tends to single out the nouns that have a high frequency of co-occurrence with the adjectives because of their high frequency in the corpus, but they may not collocate strongly with the adjectives. On the other hand, the MI measure, as explained by Church & Hanks (1990, p. 23), “compares the probability of observing x [word] and y [word] together (the joint probability) with the probabilities of observing x and y independently (chance).” The MI measure is used because compared with other association measures such as T-score which prefers function words, the MI score

tends to select content words as collocates (Church et al., 1991; Church et al., 1994), which is consistent with the purpose of this study. An MI score of 3 or higher indicates that two words tend to co-occur more frequently than would be expected by chance (Hunston, 2002). However, the MI score highlights the low frequency and exclusivity of the collocational relationship (Brezina, 2018). In other words, the noun collocates of the seven adjectives measured by the MI score tend to occur infrequently in the corpus but co-occur frequently or exclusively with the adjectives. For instance, among all the nouns modified by surprising, the word *ungallantry* has the highest MI score (14.8) because it only occurs once and co-occurs exclusively with the adjective surprising in the corpus. However, it is of little importance for the purpose of this study due to its rare occurrence in the corpus. Therefore, for the sake of the representativeness of the noun collocates measured by the MI score, I specify a threshold value of 5 for the minimum collocate frequency (i.e. the frequency of occurrence of the nouns) and collocation frequency (i.e. the frequency of co-occurrence of the nouns with the adjectives) respectively. Considering that each measure has its own strengths and shortcomings, Liu (2010) argued that the frequency measure and the MI measure can work together to help identify the typical nouns that are modified by the adjectives. Other factors, such as the collocation span (the number of words to the left and right of the node word or the search word), may also affect the identification of collocates. When using the MI statistic, I stipulate L3-R3 as the collocation span because a smaller span allows for a more focused analysis of the most immediate lexico-grammatical patterns (McEnery et al., 2019). To ensure the replicability of the results of the study, all important parameters that can impact collocate identification are reported (see Table 2) using Collocation Parameter Notation (CPN) suggested by Brezina et al. (2015).

Table 2. Collocation parameters notation (CPN) for noun collocate identification using the MI statistic*

| Statistic ID | Statistic name | Statistic cut-off value | L and R span | Minimum collocate freq. (C) | Minimum collocation freq. (NC) | Filter |
|---|----------------|-------------------------|--------------|-----------------------------|--------------------------------|--------|
| 3b | MI | 3 | L3-R3 | 5 | 5 | no |
| 3b-MI (3), L3-R3, C5-NC5; no filter applied | | | | | | |

Note. *Statistic ID refers to the number in the ID column of the original table (“b” after the Statistic ID refers to a corrected version of the MI measure, i.e. the correction for the collocation span larger than one). The filter specifies any further steps in the collocation extraction process.

The following list (see Table 3) containing the top ten noun collocates of each adjective measured by frequency and MI score (3b-MI (3), L3-R3, C5-NC5; no filter applied) was compiled. One issue worth discussing here is whether some nouns ranking in the top ten in each measure should be considered as important and relevant collocates of the adjectives to be included in the list. For example, in the case of the nouns modified by amazing, the collocates *Joseph* and *dreamcoat*, although ranking in the top 10, are excluded from the final noun collocate list because they appear mainly in the fixed phrase *Joseph and the Amazing Technicolour Dreamcoat*, which is not in line with the purpose of the current study. It is worth noting that although the two measures emphasize different aspects of the collocational relationship between the nouns and seven adjectives, there is substantial overlap between the results of the two measures: 90% overlap for staggering, 80% for shocking, 60% for startling and

stunning, 50% for astonishing, 30% for surprising and 20% for amazing. Due to significant overlap among the nouns modified by the seven adjectives, the entire list contains only 69 distinct words.

To identify the differences among the seven near-synonymous adjectives in terms of the types of nouns they modify, I categorized the nouns into six major semantic categories: (a) abstract (e.g. contrast and discovery), (b) concrete (e.g. people and color), (c) dual (e.g. thing and piece), (d) number (e.g. million and amount), (e) money (e.g. debt and cost), and (f) linguistic communication (e.g. news and story). The classification results are presented in Table 4, along with explanations at the end of the table detailing the reasons behind the classification of certain items.

Table 3. Top ten noun collocates measured by frequency and MI score

| staggering | | shocking | | startling | | astonishing | | stunning | | amazing | | surprising | |
|--------------|----------------|-----------|---------------|---------------|------------------|-------------|-----------------|----------------|-----------------|------------|-----------------|------------|----------------|
| By freq. | By MI | By freq. | By MI | By freq. | By MI | By freq. | By MI | By freq. | By MI | By freq. | By MI | By freq. | By MI |
| million 19 | billion 8 | thing 14 | waste 7.8 | contrast 16 | originality 11.3 | speed 21 | feat 10.6 | view 35 | scenery 10.8 | thing 48 | coincidence 9.3 | number 65 | agility 9.7 |
| year 11 | sum 7.6 | people 11 | news 6.9 | effect 15 | clarity 10.3 | thing 14 | speed 8.3 | effect 28 | victory 8.6 | people 29 | feat 9.0 | amount 37 | omission 8.2 |
| amount 9 | debt 7.2 | news 10 | scene 6.9 | result 14 | revelation 10.3 | amount 13 | achievement 7.9 | victory 23 | setting 7.9 | story 27 | bargain 7.8 | thing 36 | accuracy 6.6 |
| figure 7 | million 7.0 | way 9 | crime 6.7 | revelation 11 | contrast 8.6 | fact 13 | variety 6.9 | result 16 | display 7.9 | way 21 | revelation 7.6 | result 34 | discovery 6.5 |
| billion 7 | proportion 6.9 | waste 7 | behaviour 6.1 | change 9 | discovery 8.5 | success 11 | display 6.8 | goal 15 | beauty 7.8 | amount 20 | recovery 7.2 | fact 23 | finding 6.5 |
| sum 6 | amount 6.5 | story 7 | story 6.1 | clarity 9 | insight 8.5 | range 11 | success 6.5 | performance 15 | landscape 7.6 | fact 18 | discovery 6.5 | way 19 | rumor 6.2 |
| rate 6 | increase 6.2 | figure 6 | color 6.1 | beauty 8 | beauty 8.2 | variety 10 | amount 6.4 | countryside 13 | goal 7.2 | variety 16 | sight 6.5 | speed 17 | speed 5.9 |
| debt 6 | figure 5.5 | scene 6 | figure 5.3 | discovery 8 | conclusion 7.0 | piece 10 | piece 6.4 | scenery 13 | view 7.0 | time 15 | tale 6.4 | degree 14 | amount 5.9 |
| proportion 5 | rate 5.2 | color 6 | thing 5.0 | thing 6 | turn 6.5 | number 10 | progress 6.3 | color 13 | collection 6.8 | number 15 | variety 6.4 | finding 14 | lack 5.1 |
| cost 5 | cost 5.1 | book 6 | book 4.8 | fact 6 | effect 6.1 | rate 9 | degree 6.2 | display 12 | performance 6.8 | power 14 | story 6.3 | feature 12 | conclusion 4.9 |

Table 4. Type classification of the top 70 nouns modified by the seven adjectives

| Abstract | Concrete | Dual | Number | Money | Linguistic communication |
|--|--|-----------------------------|--|--------------|--|
| year way waste crime behaviour contrast effect revelation beauty discovery originality clarity insight conclusion turn change fact success feat achievement progress degree display victory performance time power coincidence recovery agility omission accuracy finding lack feature increase | people color view goal scenery countryside setting collection landscape bargain sight scene | thing* piece* result* | million amount figure billion sum rate proportion speed range variety number | debt cost | news story book tale rumor |

I then analyzed the types of noun collocates for each adjective, and the results (as presented in Table 5) reveal significant differences in the categories of nouns modified by each adjective. The types of nouns with the widest

distribution range are abstract nouns and dual nouns. These results, to some extent, shed light on the distributional patterns of these near synonyms, but they don't provide us a whole picture. For instance, the adjective startling is mostly

used to modify abstract and dual nouns, but it is not clear whether it can modify the other four types of nouns. Such information can help us gain an in-depth understanding of the meanings and usage patterns of these seven adjectives. Therefore, I searched the BNC corpus in #LancsBox X for the

co-occurrence frequencies of the seven adjectives with the 69 nouns and then computed these adjectives' total co-occurrence frequencies with each of the six types of nouns. The results are reported in Table 6.

Table 5. Types of nouns modified by each adjective

| staggering (11 different nouns) | shocking (12 different nouns) | startling (14 different nouns) | astonishing (15 different nouns) | stunning (14 different nouns) | amazing (18 different nouns) | surprising (17 different nouns) |
|------------------------------------|--|-----------------------------------|-------------------------------------|------------------------------------|--|---|
| 2 abstract 7 number 2 money | 4 abstract 3 concrete 1 number 1 dual 3 linguistic communication | 12 abstract 2 dual | 7 abstract 6 number 2 dual | 5 abstract 8 concrete 1 dual | 9 abstract 3 concrete 3 number 1 dual 2 linguistic communication | 11 abstract 3 number 2 dual 1 linguistic communication |

Table 6. Distributions of the types of nouns modified by the seven adjectives*

| Adjective | Abstract | Concrete | Dual | Number | Money | Linguistic communication | Total |
|-------------|----------|----------|------|--------|-------|--------------------------|-------|
| staggering | 47 A | 7 A | 5 A | 64 T | 11 T | 3 | 137 |
| shocking | 61 | 21 | 17 | 11 A | 0 | 24 T | 134 |
| startling | 134 T | 6 A | 23 | 20 A | 0 | 15 | 198 |
| astonishing | 136 | 27 A | 31 | 87 T | 0 | 13 | 294 |
| stunning | 128 | 112 T | 23 | 15 A | 1 | 8 A | 287 |
| amazing | 191 | 72 T | 62 T | 74 | 0 | 41 T | 440 |
| surprising | 203 | 63 | 72 T | 133 T | 1 | 21 | 493 |

Note: A cell frequency followed by the letter T means it is a "type" (discussed below) while a cell frequency followed by an A means it is an "antitype." A cell frequency followed by no letter is neither.

3.2. Semantic and Usage Differences: The Typical Types of Nouns that Each Adjective Modifies

To determine whether and how the seven adjectives' distributions among the six types of nouns differ significantly, I conducted the chi-square test and computed the expected frequency, Chi-square value, and the p-value using Python. The results are reported in Table 7. The Chi-square test showed a statistically significant difference among the seven adjectives' distributions across the six types of nouns: χ^2 (df = 42) = 672.37, $p < 0.05$. The effect size, as measured by Cramér's V, was 0.04. According to Cohen's (1988) guidelines for interpreting Cramér's V, values closer to 0.1 indicate small effects. This suggests that, while there is a statistically significant relationship between the adjectives and noun types, the practical significance of this relationship may be limited due to the small effect size.

A cell frequency that is much higher than expected is called a 'type', while a cell frequency that is much lower than expected is called an 'antitype'. On the other hand, a cell frequency that is either higher or lower than expected, but the

difference is what we would expect by chance, is considered neither a type nor an antitype. In the abstract category, startling is the only type while staggering is the only antitype. In the concrete category, two adjectives (stunning and amazing) are types, with three adjectives (staggering, startling and astonishing) being antitypes and two adjectives (shocking and surprising) being neither. In the dual category, staggering is the only antitype with only two adjectives (amazing and surprising) being types. In the number category, staggering, astonishing and surprising are types while shocking, startling and stunning are antitypes. In the money category, staggering is the only type with the other six adjectives being neither antitypes nor types. In the linguistic communication category, stunning is the only antitype and two adjectives (shocking and amazing) are types. These results show that abstract nouns are mainly modified by startling; concrete nouns are mainly modified by stunning and amazing; dual nouns are primarily modified by amazing and surprising; nouns that have to do with number are mainly modified by staggering, astonishing and surprising; nouns about money are primarily modified by staggering; nouns with regard to linguistic actions, states and processes are mainly modified by shocking and amazing.

Table 7. Chi-square test of distributions of the types of nouns modified by the seven adjectives

| Adjective | Noun Type | Freq | Exp | Chi Square | Obs Exp | P value |
|-------------|------------|------|----------|------------|---------|---------------------|
| staggering | Abstract | 47 | 59.4790 | 4.5653 | < | 3.2625073162145e-02 |
| staggering | Concrete | 7 | 19.9697 | 9.7471 | < | 1.7960487355525e-03 |
| staggering | Dual | 5 | 15.0927 | 7.3264 | < | 6.7950782176590e-03 |
| staggering | Number | 64 | 26.9202 | 66.2772 | > | 3.9176150628073e-16 |
| staggering | Money | 11 | 0.8445 | 118.8008 | > | 1.1579380539804e-27 |
| staggering | Linguistic | 3 | 8.0893 | 2.9585 | < | 8.5429140160303e-02 |
| shocking | Abstract | 61 | 58.6576 | 0.1101 | > | 7.4003112173222e-01 |
| shocking | Concrete | 21 | 19.6908 | 0.0416 | > | 8.3830566151459e-01 |
| shocking | Dual | 17 | 14.8676 | 0.2153 | > | 6.4261850047426e-01 |
| shocking | Number | 11 | 25.7056 | 10.3739 | < | 1.2780966636641e-03 |
| shocking | Money | 0 | 0.8229 | 0.1361 | < | 7.1221366476773e-01 |
| shocking | Linguistic | 24 | 8.0029 | 34.0986 | > | 5.2389932739446e-09 |
| startling | Abstract | 134 | 87.0542 | 48.9596 | > | 2.6128845865708e-12 |
| startling | Concrete | 6 | 28.0386 | 21.2054 | < | 4.1259831449275e-06 |
| startling | Dual | 23 | 21.3781 | 0.0727 | > | 7.8749347405245e-01 |
| startling | Number | 20 | 37.0162 | 9.9781 | < | 1.5841721633354e-03 |
| startling | Money | 0 | 1.1802 | 0.4338 | < | 5.1015423139078e-01 |
| startling | Linguistic | 15 | 11.4266 | 0.9655 | > | 3.2579390394007e-01 |
| astonishing | Abstract | 136 | 123.5871 | 2.2964 | > | 1.2967057748827e-01 |
| astonishing | Concrete | 27 | 40.2453 | 5.3794 | < | 2.0375886016847e-02 |
| astonishing | Dual | 31 | 30.4996 | 0.0000 | > | 9.9992711611453e-01 |
| astonishing | Number | 87 | 54.2356 | 27.1851 | > | 1.8487998309423e-07 |
| astonishing | Money | 0 | 1.6785 | 0.9556 | < | 3.2829779993377e-01 |
| astonishing | Linguistic | 13 | 16.2323 | 0.5595 | < | 4.5448160799100e-01 |
| stunning | Abstract | 128 | 120.5882 | 0.7889 | > | 3.7441951983601e-01 |
| stunning | Concrete | 112 | 40.9620 | 163.4265 | > | 2.0185965863590e-37 |
| stunning | Dual | 23 | 29.7601 | 1.6843 | < | 1.9435284522200e-01 |
| stunning | Number | 15 | 51.4182 | 35.0246 | < | 3.2556118147440e-09 |
| stunning | Money | 1 | 1.6443 | 0.0146 | < | 9.0386539539633e-01 |
| stunning | Linguistic | 8 | 15.8599 | 4.1405 | < | 4.1868703072398e-02 |
| amazing | Abstract | 191 | 177.4194 | 2.0128 | > | 1.5597585165100e-01 |
| amazing | Concrete | 72 | 57.6436 | 4.7155 | > | 2.9891630370449e-02 |
| amazing | Dual | 62 | 43.4223 | 10.2628 | > | 1.3573623330932e-03 |
| amazing | Number | 74 | 75.6748 | 0.0271 | < | 8.6924024728157e-01 |
| amazing | Money | 0 | 2.3607 | 1.8017 | < | 1.7950746056522e-01 |
| amazing | Linguistic | 41 | 23.0898 | 16.9937 | > | 3.7504555774537e-05 |
| surprising | Abstract | 203 | 195.2046 | 0.5764 | > | 4.4772922446756e-01 |
| surprising | Concrete | 63 | 62.9275 | 0.0000 | > | 1.0000000000000e+00 |
| surprising | Dual | 72 | 47.7824 | 16.3991 | > | 5.1309766376170e-05 |
| surprising | Number | 133 | 85.0073 | 40.6063 | > | 1.8620394365426e-10 |
| surprising | Money | 1 | 2.5895 | 0.5754 | < | 4.4810838412849e-01 |
| surprising | Linguistic | 21 | 25.1018 | 0.6814 | < | 4.0911171767720e-01 |

3.3. Register Distribution Patterns: Stylistic Variations

One of the significant semantic dimensions in which near-synonyms may differ is “stylistic” variation, or register variation (Edmonds & Hirst, 2002, p.109). For language learners, it is crucial to know the lexical items’ register distribution patterns to be an advanced language user. Many ESL learners tend to use formal register words in informal contexts or vice versa. Identifying the register distribution patterns of the

seven near synonyms should therefore be of great benefit to ESL learners. Thus, I first searched the BNC1994 in #LancsBox X for the frequency of each adjective across registers and then used Python to calculate the expected frequency, Chi-square value, and the p-value of each adjective in various registers. The initial results are reported in Table 8. The same annotation (“T” for “type” and “A” for “anti-type”) as in the distributions of the types of nouns modified by the seven adjectives is used here to show the distributions of the seven adjectives across registers in Table 9.

Table 8. Chi-square test of the cross-register distributions of the adjectives

| Noun | Register | Freq | Exp | Chi Square | Obs Exp | P value |
|-------------|---------------|------|----------|------------|---------|----------------------|
| staggering | spoken | 18 | 36.0408 | 9.7096 | < | 1.8330876205817e-03 |
| staggering | fiction | 120 | 78.9726 | 25.8516 | > | 3.6870338479984e-07 |
| staggering | magazine | 75 | 68.2869 | 0.6867 | > | 4.0728748280231e-01 |
| staggering | newspaper | 119 | 68.8496 | 43.5588 | > | 4.1140447204068e-11 |
| staggering | non-acad | 63 | 78.8986 | 3.7336 | < | 5.3329530108015e-02 |
| staggering | academic | 24 | 69.5520 | 35.5315 | < | 2.5096161116766e-09 |
| staggering | Miscellaneous | 126 | 117.3819 | 0.7606 | > | 3.8314435198326e-01 |
| shocking | spoken | 58 | 37.1947 | 12.6351 | > | 3.7857082992764e-04 |
| shocking | fiction | 162 | 81.5212 | 97.7771 | > | 4.6822047771232e-23 |
| shocking | magazine | 52 | 69.9968 | 5.3180 | < | 2.1106476349031e-02 |
| shocking | newspaper | 86 | 70.4960 | 3.8878 | > | 4.8637664141928e-02 |
| shocking | non-acad | 64 | 81.0837 | 4.2198 | < | 3.9954811879808e-02 |
| shocking | academic | 43 | 71.6170 | 13.4660 | < | 2.4292167614293e-04 |
| shocking | Miscellaneous | 96 | 120.2273 | 6.3446 | < | 1.1773958271204e-02 |
| startling | spoken | 8 | 41.9733 | 30.6504 | < | 3.0896195378814e-08 |
| startling | fiction | 156 | 92.4164 | 54.0431 | > | 1.9613995279807e-13 |
| startling | magazine | 92 | 79.7516 | 2.1221 | > | 1.4518871840361e-01 |
| startling | newspaper | 64 | 79.7873 | 3.5909 | < | 5.8097293127275e-02 |
| startling | non-acad | 114 | 92.4822 | 5.9935 | > | 1.4358682286651e-02 |
| startling | academic | 66 | 81.4498 | 3.3740 | < | 6.6232163113053e-02 |
| startling | Miscellaneous | 142 | 137.0678 | 0.1958 | > | 6.5817172911229e-01 |
| astonishing | spoken | 41 | 53.3862 | 3.0967 | < | 7.8449881069185e-02 |
| astonishing | fiction | 115 | 116.6000 | 0.0132 | < | 9.0843000254913e-01 |
| astonishing | magazine | 105 | 101.2089 | 0.1335 | > | 7.1479507937534e-01 |
| astonishing | newspaper | 147 | 102.0082 | 24.2498 | > | 8.4617545293144e-07 |
| astonishing | non-acad | 151 | 117.6571 | 11.7085 | > | 6.2214065296325e-04 |
| astonishing | academic | 71 | 103.2827 | 12.2339 | < | 4.6927801754069e-04 |
| astonishing | Miscellaneous | 200 | 174.7600 | 4.8652 | > | 2.7403667859944e-02 |
| stunning | spoken | 14 | 58.1156 | 38.5975 | < | 5.2086840323046e-10 |
| stunning | fiction | 91 | 126.9577 | 12.7089 | < | 3.6391761124797e-04 |
| stunning | magazine | 286 | 112.5998 | 335.3476 | > | 6.5780612767585e-75 |
| stunning | newspaper | 222 | 112.2253 | 134.2371 | > | 4.8486590388685e-31 |
| stunning | non-acad | 43 | 126.9891 | 70.4095 | < | 4.8187806476815e-17 |
| stunning | academic | 11 | 112.0454 | 113.5401 | < | 1.6432356982500e-26 |
| stunning | Miscellaneous | 247 | 191.6935 | 21.9500 | > | 2.7984846249484e-06 |
| amazing | spoken | 314 | 108.3676 | 501.2202 | > | 5.1579257461208e-111 |
| amazing | fiction | 304 | 234.6269 | 28.1834 | > | 1.1034744559887e-07 |
| amazing | magazine | 279 | 203.3766 | 37.9106 | > | 7.4060441971829e-10 |
| amazing | newspaper | 292 | 204.3354 | 50.8423 | > | 1.0009550139613e-12 |
| amazing | non-acad | 136 | 232.0694 | 54.5887 | < | 1.4858975254266e-13 |
| amazing | academic | 47 | 203.7054 | 162.9164 | < | 2.6091081210179e-37 |
| amazing | Miscellaneous | 437 | 353.0755 | 29.7971 | > | 4.7969764494383e-08 |
| surprising | spoken | 163 | 177.3441 | 1.5990 | < | 2.0604946822947e-01 |
| surprising | fiction | 387 | 391.5835 | 0.0679 | < | 7.9444521654452e-01 |
| surprising | magazine | 271 | 336.9643 | 19.8427 | < | 8.4081698540430e-06 |
| surprising | newspaper | 234 | 337.0873 | 48.6510 | < | 3.0581811755702e-12 |
| surprising | non-acad | 771 | 406.7763 | 528.3168 | > | 6.5636499027238e-117 |
| surprising | academic | 926 | 365.0151 | 1387.5952 | > | 1.0423749436183e-303 |
| surprising | Miscellaneous | 735 | 599.1959 | 52.8671 | > | 3.5690385563930e-13 |

Table 9. Distributions of the seven adjectives across registers

| | spoken | fiction | magazine | Newspaper | non-acad | academic | Miscellaneous | Total |
|-------------|--------|---------|----------|-----------|----------|----------|---------------|-------|
| staggering | 18 A | 120 T | 75 | 119 T | 63 | 24 A | 126 | 545 |
| shocking | 58 T | 162 T | 52 A | 86 T | 64 A | 43 A | 96 A | 561 |
| startling | 8 A | 156 T | 92 | 64 | 114 T | 66 | 142 | 642 |
| astonishing | 41 | 115 | 105 | 147 T | 151 T | 71 A | 200 T | 830 |
| stunning | 14 A | 91 A | 286 T | 222 T | 43 A | 11 A | 247 T | 914 |
| amazing | 314 T | 304 T | 279 T | 292 T | 136 A | 47 A | 437 T | 1809 |
| surprising | 163 | 387 | 271 A | 234 A | 771 T | 926 T | 735 T | 3487 |

Note: A cell frequency followed by the letter T means it is a “type” (discussed above) while a cell frequency followed by an A means it is an “antitype.” A cell frequency followed by no letter is neither.

Considering that spoken language and academic writing represent the two extremes of the formality scale shown by the seven registers with spoken being the least formal and

academic writing being the most formal, it is sensible to focus on how the seven adjectives distribute in these two registers. Based on the detailed results of the chi-square test reported in

Table 9, shocking and amazing appear to be the most informal because they both are types in the spoken register and anti-types in the academic writing register, while surprising appears to be the most formal because it is the only type in the academic writing register. Staggering and stunning are in the middle as they are anti-types in both registers. Startling is

more formal than astonishing because the former is the antitype in the spoken register while the latter is the anti-type in the academic writing register. Figure 1 below shows the seven adjectives' formality scale (from less formal to more formal).



shocking/amazing<staggering/stunning<astonishing<startling<surprising

Figure 1. The formality scale of the seven adjectives

4. Conclusion

Through a corpus-based behavioral profile analysis of seven English synonymous adjectives, this study has identified several usage and stylistic variations among them. First, each of the seven adjectives tends to modify different types of nouns. Second, the register distribution patterns of the seven adjectives are different. Thus, they display different levels of formality (as illustrated in Figure 1). The results of the study remedied the shortcomings of the traditional references that were insufficiently descriptive of synonymous adjectives and provided useful information for ESL teachers and learners. Furthermore, the study has proved that corpus-based BP research is a suitable and effective approach for investigating a set of synonymous adjectives. Meanwhile, this study also contributes to the recent literature on the semantic and usage patterns of specific sets of near synonyms. However, this study did not explore the potential influence of the syntactic position on these adjectives' usage. Further research can examine whether the syntactic position (i.e. attributive or predicative or both) has an influence on the usage patterns of the seven adjectives.

References

- [1] Bolinger, D. L. (1968). Entailment and the meaning of structures. *Glossa*, 2, 119–127.
- [2] Brezina, V. (2018). Statistics in Corpus Linguistics: A Practical guide. <https://eprints.lancs.ac.uk/id/eprint/131100/>
- [3] Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- [4] Cruse, D. A. (2000). *Meaning in language: An introduction to semantics and pragmatics*. Oxford University Press.
- [5] Church, K. W., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: Exploiting on-line resources to build a lexicon* (pp. 115–164). Lawrence Erlbaum.
- [6] Church, K. W., Gale, W., Hanks, P., Hindle, D., & Moon, R. (1994). Lexical substitutability. In B. T. S. Atkins & A. Zampolli (Eds.), *Computational approaches to the lexicon* (pp. 153–177). Oxford University Press.
- [7] DiMarco, C., Hirst, G., & Stede, M. (1993). The semantic and stylistic differentiation of synonyms and near-synonyms. In *Building lexicons for machine translation volume*. AAAI.
- [8] Divjak, D. (2006). Ways of intending: Delineating and structuring near-synonyms. In De Gruyter eBooks (pp. 19–56). <https://doi.org/10.1515/9783110197709.19>
- [9] Divjak, D., & Gries, S. T. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2(1). <https://doi.org/10.1515/cllt.2006.002>
- [10] Divjak, D. (2010). Structuring the lexicon: A clustered model for near-synonymy. <http://ci.nii.ac.jp/ncid/BB04508631>
- [11] De Saussure, F. (2011). *Course in General Linguistics*. Columbia University Press.
- [12] Edmonds, P., & Hirst, G. (2002). Near-Synonymy and lexical choice. *Computational Linguistics*, 28(2), 105–144. <https://doi.org/10.1162/089120102760173625>
- [13] Firth, J. R. (1957). A synopsis of linguistic theory 1930–55. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth 1952–59*. Longmans. (Original work published 1957)
- [14] Gries, S. Th. (2001). A corpus-linguistic analysis of English -ic vs -ical adjectives. *ICAME Journal*, 25, 65–108.
- [15] Gries, S. Th. (2010). Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, 5(3), 323–346. <https://doi.org/10.1075/ml.5.3.04gri>
- [16] Gries, S. T., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34, 121–150.
- [17] Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. Bazell, J. Catford, M. Halliday, & R. Robins (Eds.), *In memory of J. R. Firth* (pp. 148–162). Longman.
- [18] Harris, Z. S. (1970). *Papers in Structural and Transformational Linguistics*. In Springer eBooks. <https://doi.org/10.1007/978-94-017-6059-1>
- [19] Hanks, P. (1996). Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1(1), 75–98. <https://doi.org/10.1075/ijcl.1.1.06han>
- [20] Kaunisto, M. (1999). Electric/electrical and classic/classical: Variation between the suffixes -ic and -ical. *English Studies*, 4, 343–370.
- [21] Liu, D. (2010). Is it a chief, main, major, primary, or principal concern? *International Journal of Corpus Linguistics*, 15(1), 56–87. <https://doi.org/10.1075/ijcl.15.1.03liu>
- [22] Saussure, F. de. (1983). *Course in general linguistics* (R. Harris, Trans.). G. Duckworth. (Original work published 1916)
- [23] Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- [24] Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.
- [25] Taylor, J. R. (2002). Near synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Sciences*, 25(3), 263–284. [https://doi.org/10.1016/s0388-0001\(02\)00018-9](https://doi.org/10.1016/s0388-0001(02)00018-9)

[26] Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of

business English. *English for Specific Purposes*, 30(2), 101–112. <https://doi.org/10.1016/j.esp.2010.12.003>