A Corpus Study of Near-Synonyms and Their Lexicographic Representation

Sara Sowers-Wills

University of Minnesota Duluth

Abstract Working from corpus data collected for near-synonymous adjectives, this paper argues that the subtle differences found in the near-synonymous relationships between the senses of lexical units are relevant for the structure of the lexicon and also have implications for lexicography. Whether there is definable nuance in meaning that distinguishes the use of one near-synonymous adjective over another is an important question. If usage trends found in corpus data reveal subtle but definable semantic nuance between near-synonyms, then this fine-grained distinction should be reflected in dictionaries—particularly learner's dictionaries since language learners depend on their content to achieve native-sounding language. The present study examines the nature of near-synonymy in English by analyzing semantic relationships in two pairs of near-synonymous adjectives (i.e., *big*/*large* and *enormous*/*huge*) and the nouns with which they frequently collocate. Data were collected from *The Corpus of Contemporary American English*, and WordNet Search 3.1 online was used to sort the nouns modified by each of the near-synonymous adjectives into semantic categories. Edmonds and Hirst's (2002) cluster model of lexical knowledge provides suitable representation of the fine-grained aspects of meaning and behavior that distinguish near-synonyms. The model facilitates interpretation of the data presented here, which may benefit the representation of near-synonyms in English learner's dictionaries in the future.

Keywords near-synonymy; collocation; lexicography

> ‘And things which have a relation are near enough to have something in common, but not near enough to be one and the same, are they not?’
> —Gerard Manley Hopkins, “On The Origin Of Beauty: A Platonic Dialogue”

1. Introduction

The epigraph from Gerard Manley Hopkins inadvertently but succinctly describes near-synonymy. One can go to an English-language thesaurus, look up *big*, and find several contenders: *large*, *huge*, *enormous*, *immense*, *massive*, *gigantic*, and so on. In fact, in *Roget’s International Thesaurus, Fifth Edition* (1992), one finds the adjective *big* under the category of size, with such synonyms as *large*, *sizable*, *big*, *great*, *grand* grouped together and *huge*, *immense*, *vast*, and *enormous* grouped together. These words are all similar in meaning—here concerning the sense of more than average size, but for one reason or another, one of these words is selected over the others to modify a particular noun. These words are near-synonyms, close enough to have something in common and some degree of interchangeability, yet distant enough not to be the same. It has been conjectured that this kind of relationship does not impact the structure of the lexicon. For example, Lipka (1990) references the semanticist Lyons (1968:452f.) who wrote that synonymy is “fairly irrelevant for the structure of the lexicon of a language”. This study argues that the subtle differences found in the near-synonymous relationships between the senses of lexical units are, in fact, relevant for gaining insight into the structure of the lexicon and also have implications for lexicography. Working from corpus data collected for near-synonymous adjectives, the present work examines whether usage trends for this phenomenon are accurately reflected in English learner’s dictionaries.

If someone were to ask, “What’s the large deal?”, rather than the idiomatic “What’s the big deal?”, this person would quickly be pegged as a non-native speaker of English. Nuance like this comes naturally to a native speaker but can be troublesome to a non-native learner (Laufer, 1990; Martin, 1985). While *large* and *big* are generally seen as synonymous, something makes the use of *large* in this context sound strange and makes *big* just sound right. “What’s the huge deal?” or “What’s the enormous deal?” would sound just as strange as when *large* is used instead of *big*. One can easily find in everyday spoken English the phrase, “It isn’t a huge deal”, but rarely an *enormous* deal. Certainly, some adjective-noun combinations, like *big deal*, have assumed collocational strength so to be inseparable and attain meaning only in their co-occurrence. This particular collocation is represented as an entry in some learner’s dictionaries. After all, what *deal* means here is definable but much more meaningful in its collocational context. In contrast, other adjective-noun collocations tend only to co-occur but do not necessarily attain their meaning as a unit. For example, is it possible to determine which of a set of near-synonymous adjectives denoting bigness is typically used to modify a building, a person, an amount, or any abstract entity?

Whether there might be some definable nuances of meaning that distinguish the use of one of these near-synonymous adjectives over another is an important question. If there are not such definable nuances, the selection is random. If the selection is random, there should be no patterns to be found in a corpus, which contribute to measures of predictability, and the reverse should be true if the selection is not random. If such usage trends are to be found in corpus data, however, it is important to home in on fine-grained distinctions between near-synonyms in order to accurately reflect meaning and usage in

learner's dictionaries (Biber et al., 1994; Verlinde & Selva, 2009). The present study examines the nature of near-synonymy by analyzing semantic relationships in four near-synonymous adjectives and the nouns with which they tend to collocate.

Getting to the heart of the possible scenarios posed above inevitably leads to deeper consideration of the nature of synonymy or, more accurately, near-synonymy. Linguists generally agree that absolute synonymy is for the most part impossible, or at least very rare (Cruse, 1986; Divjak, 2010; Stubbs, 2001). The only places in which we might find absolute synonymy are in technical terms or dialectal variation. For example, *groundhog* and *woodchuck* refer to the same animal, with no difference in meaning. One could argue that these are absolute synonyms, but one could also argue that whichever of the two words one uses reflects a regionalism. Cruse (1986: 268) states that two lexical units could be considered to be absolute synonyms "if and only if all their contextual relations….were identical". In contrast, Taylor (2002: 266) discusses the distinction between the pair *big* and *large* and the pair *warm* and *hot*. The latter pair occupies different places on a scale, but *big* and *large* are not contrastive in this way; neither are they interchangeable. They have subtly different distributions, which points to differences in meaning (Taylor, 2002). An adjective in particular can be complex because meaning can shift depending on the noun it modifies (Murphy, 2010: 222-223). Near-synonyms can differ in many different ways, including denotation, connotation, style, register, or emphasis (Cruse, 1986; Edmonds, 1999), and contextual relations are important to understanding the construction of meaning (Evans & Green, 2006). Cruse (1986) describes near-synonyms as identical in terms of their "central" traits but different in their "peripheral" traits, which is a suitable place to begin, but it is not clear how to discern

which traits are central and peripheral given the dynamic nature of adjectives. Even native speakers struggle to point to the fine distinctions between near-synonymous adjectives, so a finer-grained approach to these concepts is needed to improve the quality of learner's dictionaries for non-native speakers.

Most research addressing near-synonymy has focused on verbs or nouns (e.g., Divjak, 2010; Divjak & Gries, 2008; Edmonds & Hirst, 2002; Xiao & McEnery, 2006). In a corpus-based study, Liu (2010) subjected a set of five near-synonymous English adjectives to behavioral profile analysis (see Gries, 2010), finding fine-grained nuances in meaning. Gries and Otani (2010) similarly conducted a corpus-based behavioral profile study of near synonymy-antonymy relations in the semantic field of *SIZE* in English and found nuances of meaning associated with the target words and their canonical antonyms. Edmonds and Hirst (2002) depart from Cruse's (1986) description of near-synonymy, presenting a clustered model of lexical knowledge in which a group of near-synonyms clusters together under a common coarse-grained meaning. This model provides a way to represent finer-grained aspects of meaning that distinguish the near-synonyms of a cluster (Edmonds and Hirst, 2002: 106), dispensing with the central-peripheral dichotomy (Cruse, 1986). Their model offers a refinement of the representational organization and behavior of near-synonyms, facilitating interpretation of the data presented here, which stands to benefit the representation of near-synonyms in English learner's dictionaries.

The present research builds on the author's previously unpublished work examining five near-synonymous adjectives and a set of nouns with which they frequently collocate, using corpus data. High-frequency nouns collocating with each of the five near-synonymous adjectives—*big*, *large*, *enormous*, *huge*, and *fat*—were

collected from *The Corpus of Contemporary American English (COCA)* (Davies 2008-). Containing 450-million words, COCA is a large and balanced part-of-speech (POS)-tagged corpus of American English. It was selected as the most suitable corpus because it is set up for the kind of search parameters needed for this project—namely, collocations restricted to immediately collocating words, which can be searched by genre and sorted by frequency and mutual information (MI) score. In addition, results can be grouped by lemma so as not to exclude varying relevant forms of a search query. The nouns found to frequently collocate with the adjectives were semantically classified using the WordNet database in order to achieve both objectivity and consistency in classification. It was not always clear within the parameters of WordNet how to categorize a particular noun in the way it was used in the corpus data, necessitating some subjectivity and potentially some degree of inaccuracy.

Patterns to be discussed below emerged from the corpus results, and implications of the relationships found between the selected near-synonymous words point to at least two important areas of investigation—lexicography and lexical semantics. With respect to the former, learner's dictionaries are of particular interest, where the nuances of usage of near-synonymous words can be troublesome for learners of English to grasp, rendering their representation in learner's dictionaries quite important. In this paper, entries for four adjectives are examined in four of the five major English learner's dictionaries to see whether the most significant corpus patterns are reflected in respective dictionary entries. The adjective *fat* was excluded in order to focus on more fine-grained relationships between two pairs of near-synonyms: *big*/*large* and *huge*/*enormous.* With respect to lexical semantics, collocational patterns for near-synonymous adjectives can provide

insight into lexical relations and the way that representations for these concepts are organized in the lexicon. This research also contributes to choices about how lexical items are represented in dictionaries. Section 2 briefly describes the methodology and findings that prompted the present research. Section 3 discusses the representation of the selected near-synonyms in four learner's dictionaries, allowing for a closer examination of the nature of near-synonymy in Section 4. Section 5 ends the paper with a discussion of some conclusions that can be drawn about near-synonymy and how it is represented in learner's dictionaries, along with suggestions for future work.

2. A brief description of the corpus methodology and findings

This section describes the previous corpus research and the findings upon which the present paper builds. Because definitions for what a collocation is range from loose to more restricted, how the term is used in the present research needs to be clarified. While frequency statistics were used to measure collocational strength, a somewhat loose definition of collocation is assumed, following Firth (1968: 181, cited in Xiao & McEnery 2006), in that "collocations of a given word are statements of the habitual or customary places of that word", rather than Hoey (1991: 6-7), who ascribes to collocation the property of a given word occurring with another "with greater than random probability in its (textual) context".

It is precisely this looser definition that is of interest here: that is, the strong tendency to choose one near-synonym over another, not because a particular noun and adjective have become a semantic unit in combination but because something just seems right about it to the speaker. Intuition implicitly embedded in the native speaker's

knowledge is involved in this sort of decision. Notably, intuition as such is typically not available to non-native speakers, justifying the inclusion of collocational and subtly nuanced meaning in learner's dictionaries. The question asked here is whether the semantic properties of a set of near-synonyms are predictable.

Using *COCA* (Davies 2008-), the 30 most frequent nouns with which five near-synonymous adjectives collocate were collected. The choice to examine the 30 most frequent collocating nouns was motivated by the exploratory nature of this work. Initially, only the 10 most frequent collocating nouns were examined, but this did not reveal clear enough patterns. Extending the analysis to include the 30 most frequently occurring nouns in collocation with the target adjectives provided more data from which patterns could more clearly emerge, while also constraining the data set in order to focus the analysis. As an exploratory study, this work is meant to serve as a first step to future related work, in which the selection of the data set should be more systematic in view of the overall distribution of nouns for each adjective. The adjectives examined include *big*, *large*, *huge*, *enormous*, and *fat*. These were selected for their commonness in the language in general, on English learner websites, and because they are frequently listed as being synonymous with one another. *Big* was selected as the basic adjective, which may play a role in defining its near-synonyms. *Large* was selected because of its close semantic relationship with *big*; that is, they are both relatively neutral. *Huge* and *enormous* were selected because they were predicted to be less frequent than *big* and *large* as well as less neutral in meaning, indicating a greater degree of bigness. Finally, *fat* was selected because it was deemed the least semantically neutral, which could be useful for contrastive purposes. Table 1 presents the representation of each of the

adjectives within COCA: the number of collocating nouns with more than one instance in the corpus are shown in the middle column, and the overall frequency of the adjective is shown in the right-hand column. One might expect that *big*, with its frequency higher than the other adjectives and larger number of collocating nouns, to collocate with nouns of a greater variety of semantic categories—and the converse for enormous and fat, which have a much lower number of collocating nouns and much lower overall frequencies. The task of this research is to determine the semantic categories in the most frequent collocations and determine whether or not these categories are reflected in English learner's dictionaries.

Table 1. Total number of nouns collocating with each target adjective (of which there is more than once instance in the corpus.

	# collocating nouns	**overall frequency of adj.**
big	3209	239823
large	2800	142812
huge	1887	57277
enormous	956	21778
fat	540	10228

COCA was determined to be an appropriate search tool for the project because of its size, its balance and representativeness, its capacity for clearly defining a collocational search by part of speech, and its ability to present frequency and statistical data in search results. It is also large, genre-balanced, and up-to-date, which in theory should reap results that are representative and accurate in terms of American English in use.

In order to limit the data in a way that could be consistently examined, the search query was set to find the target adjective plus an immediately following noun, sorting by frequency and requiring a mutual information (MI) score of 3 or above to ensure collocational strength, and using only adjectives of high relative frequency. Although

searching for nouns immediately following the adjective excludes many occurrences of collocations that could be informative for the task at hand, this method was selected as an appropriate place to begin in order to keep the data clean, so that in continuing the project there will be distinct data for the selected adjectives collocating with immediately following nouns and data for nouns more distantly collocating with the selected adjectives. More specifically, while adjectives are also commonly found in sequence with other adjectives (i.e., "the big, fat cat"[1]) or within a relative clause (i.e., "a boat which is big"), the present work serves as a preliminary study focusing on only one collocational configuration—that concerning a target adjective immediately followed by a noun. The intention was to collect nouns that frequently and characteristically co-occur with the target adjectives. The nouns were grouped by lemma in order not to exclude, for example, plural forms that would be as relevant as their singular counterparts. A separate search was conducted for each of the three genres investigated: newspaper, fiction, and academic, although the question of genre effect will not be addressed here.

Upon collecting collocation data, WordNet Search 3.1 online was used to categorize the semantic domains for each noun. This database was chosen in order to approach semantic classification objectively and systematically. Each noun was searched on the WordNet Search 3.1 website, and then its semantic relations were examined, specifically focusing on inherited hypernyms (i.e., the semantic relation in which a hypernym, a word with a more broad meaning, is superordinate to more specific, semantically related words). Examining inherited hypernym hierarchies allowed for the consistent classification of nouns and for the determination of multiple possible groups

[1] While sequential adjectives preceding a noun are relatively common, this pattern did not occur among the most frequent occurrences of the adjective-noun collocations, which were examined in the present work.

where meaning might intersect or be important for discerning collocational tendencies. It became important while categorizing nouns according to their semantic qualities to return to the contexts in which some of the tokens were found. Many of the tokens bore variously nuanced meanings. Special attention was given to the fact that many words take on nuanced meaning when in collocation with the target adjectives. The nouns in a few collocations were not captured exactly in WordNet (e.g., 'big money'). In order to achieve consistency and as much objectivity as possible, the closest inherited hypernyms were used. In some cases, subjective judgment was required in determining the use of words in their contexts, despite the use of WordNet, in order to decide which hypernym hierarchy to use. This was, furthermore, the case because the nouns in the WordNet system do not all have the same hierarchical structure (e.g., some have more levels than others).

In the original data collection, nouns were classified between abstract and physical, animate and inanimate, basic/general category, and specific category. For example, *man* is classified as physical, animate, living thing, and human, respectively. Not all nouns could be so neatly classified. Constant revision of the classification of the list of nouns to be sure that the nouns had been classified according to terms that, where appropriate, matched and could be assessed statistically as groups. While most nouns could be given a single classification, some held varying shades of meaning (e.g., 'energy', as a physical process or as an industry, for example). Classification of these nouns was challenging; the intention was to maintain balance between keeping classification general enough to encompass the varying shades of meaning yet still specific enough to warrant the classification scheme designated for the rest of the nouns.

For example, among the data for *big*, *picture* was an abstract noun, but among the data for *large* it was a physical noun. Differences in usage, like this, were taken into consideration in the semantic classification of nouns.

Collecting and categorizing the data in this way using WordNet was useful to a limited extent for locating semantic trends of a more specific nature, which would be difficult to determine using only intuition. The results of this corpus study met the general hypotheses about how each of the adjectives would behave in terms of relative distinction from one another and differences in occurrence depending on the genre of the text in which the collocation appeared. For example, when *fat* is used to describe animate nouns, it is seen in fiction (84.53% of occurrences) far more than in newspaper (18.21%) or academic writing (22.11%) (see Table 2). This is expected since *fat* tends to have a pejorative connotation, which is acceptable in fiction but not in more formal writing. In contrast, *fat* describes inanimate nouns in newspaper (81.79%) and academic writing (77.89%) much more often than in fiction (15.47%) (see Table 2). This is also expected for the same reason.

Table 2. Distribution of the adjective *fat* used to modify animate vs. inanimate nouns across the genres fiction, newspaper, and academic writing.

	animate	**inanimate**
fiction	84.53%	15.47%
newspaper	18.21%	81.79%
academic writing	22.11%	77.89%

Furthermore, in a majority of instances, *enormous* (71.26%) and *huge* (51.77%) collocate with nouns that are categorized as attributes by WordNet. In the case of *enormous*, this includes nouns that most often subcategorize into states, amounts, or assets, like *problem*, *number*, or *resource*, respectively. In the case of *huge*, this includes nouns that most often subcategorize into states, amounts, or assets, like *problem*, *number*,

or *advantage*. As one can see, there is a close similarity between *enormous* and *huge* in this small sample. The near-synonymous adjectives *big* and *large* are both frequent and relatively neutral in meaning. It is interesting to see, then, that while a pattern in the semantic domains of nouns collocating with *big*, with *large* is more subtle, there is nonetheless a detectable trend toward nouns in the attribute (e.g., magnitude, amount, degree) and object (e.g., body, area, structure, culinary) categories. This may be because *big* is more frequent than *large*, and more frequent words acquire more varied usage in terms of the contexts and collocations in which they are found. This is interesting because these two adjectives seem to be very close in meaning yet find themselves in different collocational environments.

The present study follows up on these findings. The next section addresses how *big*, *large*, *huge*, and *enormous* are treated in four of the five major learner's dictionaries, specifically comparing the pair *big* and *large* and the pair *huge* and *enormous*. *Fat* is excluded in order to closely examine the more subtle differences in these two pairs of synonyms.

3. Lexicographic representation

Four learner's dictionaries were used in this research: The *Cambridge Advanced Learner's Dictionary, Third edition* (*CALD*), the *Longman Dictionary of Contemporary American English* (*LDOCE*), the *COBUILD Advanced British English Learner's Dictionary, Second edition* (*COBUILD*), and the *Oxford Advanced Learner's Dictionary,*

Seventh edition (*OALD*). These learner's dictionaries are based on large corpora comprised of written and spoken text spanning several genres and present both British and American English usage, differing to some degree in their representation of each. It may be argued that a more direct assessment of the lexicographic representation of American English usage, as found in the American English data on the COCA database, will be attained by examining learner's dictionaries of American English only. However, the exploratory nature of this study prompted an examination of four of the five major English learner's dictionaries available in order to gain a broad, foundational view of lexicographic representation for English in connection with trends found in a natural-language corpus. In the Appendix following the references are tables displaying information derived from the lexicographic representation of each of the four adjectives examined, including definitions and collocational examples.

First, let us address the corpus representation of collocational trends for *huge* and *enormous* in relationship with how these words are represented in English learner's dictionaries. The hierarchic semantic categories taken from WordNet were either too broad, too specific, or too disparate for the purposes here, so the nouns that collocate with the four adjectives based on each other were re-classified. This proved easier to work with and more sensible for the current analysis than the categories provided by WordNet. Definitions presented in English-language dictionaries and the researcher's informed, native knowledge of the language, guided by the original semantic classification based on the hierarchies found on WordNet, were used to re-categorize nouns. The nouns that collocate with *huge* and *enormous* were sorted into the following categories: amount (A), challenge (C), neutral effect/outcome (E), capacity used to make something happen (F),

group/mass (G), negative effect or outcome (N), positive effect or outcome (P), and range or distance from one point to another (either abstract or concrete; R). Each noun was sorted into only one category. One can easily imagine any of these nouns falling into more than of the categories devised. The context in which each token occurred was consulted and guided the re-classification process. Each noun was classified according to its use in context, and dictionary definitions were used to guide the classification of each of these words into the same category. For example, *step* and *swath* are both categorized as range (R). Relevant dictionary definitions for each of these words—in consideration of their usage in context—include the notion of a defined "space" or a measurable "distance". These definitions connect the two words under the category of "range" employed in this study. In future work, building on this preliminary study, a reliability measure could help to ensure accuracy in the semantic classification of each noun.

Table 3 shows how each of the nouns has been re-classified. Below, Table 4 shows the numeric breakdown of each of the nouns into the semantic domains that emerged, and Figure 1 offers a visual representation of the distribution of the nouns across the semantic categories. There are other ways in which these lexical items might be semantically categorized. For example, it might be viewed as questionable to have positive and negative categories in addition to the other categories, where there is bound to be some overlap. The positive and negative categories were chosen because many of the nouns in the list very clearly have either positive or negative connotation, where others could go either way and are more accurately classified within a more specific semantic domain. The near-synonymous adjectives examined here may show a tendency

to collocate with nouns of positive or negative connotation, a matter that warrants further exploration.

Table 3. Semantic re-classification of the 30 most frequent nouns collocating with *huge* and *enormous*.

	huge		**enormous**	
1	Benefit	P	Problem	N
2	Problem	N	Resource	F
3	Opportunity	P	Benefit	P
4	Potential	F	Opportunity	P
5	Size	A	Number	A
6	Cost	A	Power	F
7	Advantage	P	Size	A
8	Number	A	Weight	A
9	Investment	F	Cost	A
10	Volume	A	Investment	F
11	Gap	R	Challenge	C
12	Profit	P	Potential	F
13	Amount	A	Profit	A
14	Sum	A	Pressure	F
15	Quantity	A	Sum	A
16	Swath	R	Popularity	P
17	Range	R	Quantity	A
18	Population	G	Amount	A
19	Market	G	Range	R
20	Collection	G	Increase	P
21	Crowd	G	Growth	P
22	Pile	G	Energy	F
23	Body	G	Change	E
24	Increase	P	Effort	F
25	Step	R	Risk	C
26	Difference	R	Difference	R
27	Impact	E	Success	P
28	Success	P	Influence	F
29	Expansion	P	Difficulty	N
30	Loss	N	Impact	E

P=positive; N=negative; G=group/mass; A=amount; R=range (abstract or concrete); E=effect/outcome; F=capacity; C=challenge

For the majority of semantic domains, *huge* and *enormous* have similar collocational tendencies. There are three points, however, on which collocational tendencies diverge. Among the 30 most frequently collocating nouns, where *huge* collocates with nouns that can be classified into “group or mass”, *enormous* does not. In

addition, where *enormous* collocates with nouns suggestive of "capacity" in 27% of the most frequent nouns, for *huge* this trend is only at 7%. While the data sample is small and the numbers reported must be held in view of this, one nevertheless expects to see definitions and examples representative of these trends in dictionaries, especially in learner's dictionaries, where nuances in meaning are particularly important for inclusion.

Table 4 and Figure 1 each show the distribution of the 30 most frequently collocating nouns with the adjectives *huge* and *enormous* in terms of types. Analyzing the collocation data in this way, we can point to specific words and their associated semantic categories that tend to collocate with either of these adjectives, or in some cases both adjectives. Examining token frequencies, however, can also be informative for gaining insight into how deeply entrenched an association between the target adjectives and the semantic categories of collocating nouns—and specific collocations—may be. With this in view, Table 5 and Figure 2 display the distribution of the semantic domains for nouns that tend to collocate with *huge* and *enormous* in terms of token frequency. Overall, the distribution of collocating nouns across the semantic categories is similar between type and token frequency, although the balance shifts to more heavily weight "amount" (A) nouns collocating with each *huge* and *enormous*. Additionally, less frequent occurrence of "positive" (P) nouns collocating with *enormous* is seen when looking at token frequency (11%) in comparison with type frequency (20%). Commentary on these points is kept brief and observational here in order to maintain focus, but differences between type and token frequency stand as an important issue that in future research could shed light on collocational tendencies across near-synonyms.

Table 4. A numeric breakdown of the semantic domains into which each of the nouns has been re-classified (type frequency). Notable differences between the two adjectives are in bold.

	huge	**enormous**
A	6 (20%)	8 (27%)
C	1 (3%)	2 (7%)
E	2 (7%)	2 (7%)
F	1 (3%)	**8** (27%)
G	**4** (13%)	0 (0%)
M	2 (7%)	0 (0%)
N	4 (13%)	2 7%)
P	4 (13%)	6 (20%)
R	3 (10%)	2 (7%)
T	3 (10%)	0 (0%)

P=positive; N=negative; G=group/mass; A=amount; R=range (abstract or concrete); E=effect/outcome; F=capacity; C=challenge; T=natural entity; M=miscellaneous. Percentages are relative to the 30 most frequently collocating nouns with each adjective, given for quick numeric perspective; total percentage may be higher than 100 due to rounding errors.

Figure 1. Distribution of semantic domains for nouns collocating with *huge* and *enormous* (type frequency).

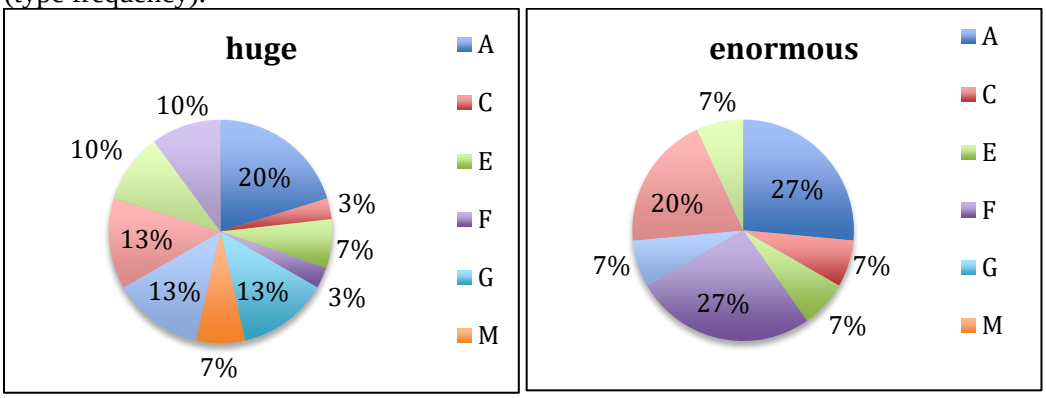


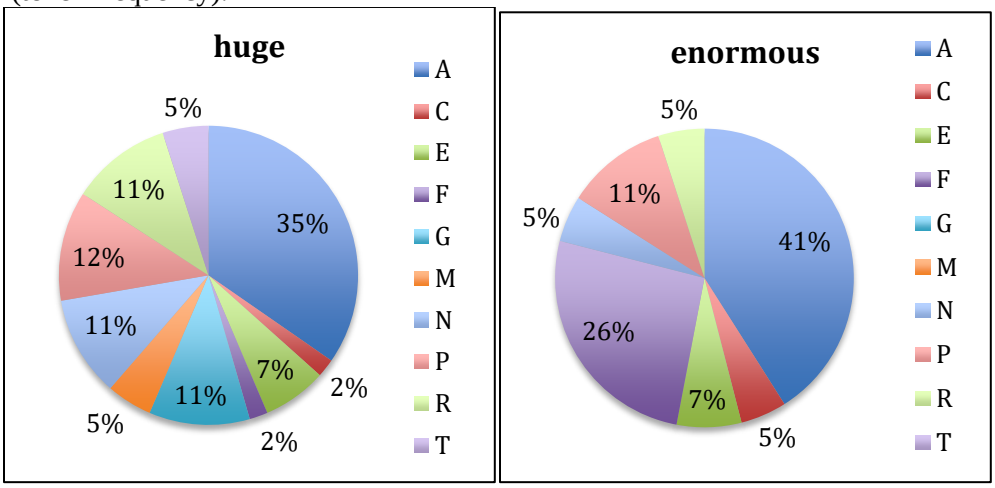
Table 5. A numeric breakdown of the semantic domains into which each of the nouns has been re-classified (token frequency). Notable differences between the two adjectives are in bold.

	huge	**enormous**
A	1854 (35%)	1256 (41%)
C	92 (2%)	155 (5%)
E	354 (7%)	223 (7%)
F	99 (2%)	**794 (26%)**
G	**572** (11%)	**0** (0%)
M	250 (5%)	0 (0%)
N	574 (11%)	138 (5%)

P	658 (12%)	332 (11%)
R	**573** (11%)	**155** (5%)
T	292 (5%)	0 (0%)

P=positive; N=negative; G=group/mass; A=amount; R=range (abstract or concrete); E=effect/outcome; F=capacity; C=challenge; T=natural entity; M=miscellaneous. Percentages are relative to the 30 most frequently collocating nouns with each adjective, given for quick numeric perspective; total percentage may be higher than 100 due to rounding errors.

Figure 2. Distribution of semantic domains for nouns collocating with *huge* and *enormous* (token frequency).



Each of the four dictionaries consulted broadly represents each adjective as being marked by great size or degree; some specify further by way of exemplification and others do not. Further specification is never explicit. Table 6 shows the lexicographic results for *huge* and *enormous*. The *CALD* does not provide definitions that offer clear distinction between the two adjectives. Each is defined as "extremely large", and the examples of collocating nouns given for each are comparable (see Appendix). The *LDOCE* is not much more helpful. While the definition for *huge* could be said to allude to the idea of range, this idea appears as only a weak tendency in the corpus data. Furthermore, the definition for *huge* includes "capacity" as something that *huge* can express in an excessive degree. The corpus data show that *enormous* has some, albeit limited, tendency to collocate with nouns of the "capacity" category but not *huge*. The

COBUILD is somewhat more helpful, with the third definition for *huge* as follows: "Something that is huge exists or happens on a very large scale, and involves a lot of different people or things". This is not reflected in the definition for *enormous*, which is representative of the pattern found in *COCA*, in that *huge* shows a collocational trend with groups or masses of things or people and *enormous* does not. The *Oxford Advanced Learner's Dictionary* (*OALD*) also provides *crowd* as a collocational example with *huge* (i.e., *huge crowd*) but nothing that could be so classified with *enormous*. Furthermore, the *COBUILD* gives as an example for *enormous* the noun *capacity*, which belongs to the "capacity" category.

Table 6. Lexicographic results for the assessment of collocational patterns.

		HUGE	ENORMOUS
CALD	CAPACITY		
	CROWD		
COBUILD	CAPACITY		√
	CROWD	√	
OALD	CAPACITY		
	CROWD	√	
LDOCE	CAPACITY	X	
	CROWD		

√= semantic category for nouns found in corpus data is corroborated in the dictionary consulted
X = semantic category for nouns found in corpus data is found in the dictionary consulted when it should not be; no mark indicates that there is no representation of the semantic category found in the dictionary consulted

The findings from *COCA* show a tendency for *enormous*, but not for *huge*, to collocate with nouns in the semantic domain of range. The trend for *huge*, and less so *enormous*, to collocate with nouns of range or distance from one point to another is not present in the entries of any of the four dictionaries. However, as noted, this was only a weak trend and so may not warrant exemplification. In light of lexicographic entries

indicating “more than average size” for the full set of near-synonymous adjectives, additional corpus research could help shed light on this observation. In all cases where the findings from *COCA* are reflected in the dictionaries consulted in this research, the semantic domain of frequently collocating nouns is not noted explicitly but representative examples are present for the attentive reader to use for inference.

Now let us look at the corpus representation of collocational trends for *big* and *large* in relationship with how these words are represented in English learner’s dictionaries. Table 7 displays the semantic re-classification of the 30 most frequent nouns collocating with both *big* and *large*. For many semantic domains, *big* and *large* have similar collocational tendencies. There are two points, however, where collocational tendencies distinctly diverge. Where *big* collocates with nouns that can be classified into “issue or situation”, *large* does not at all. In addition, although less significantly, *large* shows a tendency to collocate with nouns classified as “part, portion, or measurement”, and *big* shows less of a tendency toward this classification. Again, it is acknowledged that the data sample is small and that the numbers reported must be held in view of this. One nevertheless expects to see definitions and examples in some way representative of these trends in dictionaries, especially in learner’s dictionaries.

Table 7. Semantic re-classification of the 30 most frequent nouns collocating with *big* and *large*.

	Big		**Large**	
1	Picture	I	Volume	A
2	Thing	I	Majority	P
3	Problem	I	Extent	R
4	Question	I	Number	A
5	Smile	M	Amount	A
6	News	I	Sum	A
7	Business	G	Quantity	A
8	League	G	Company	G

9	Screen	G	Group	G
10	Company	G	Population	G
11	Deal	I	Crowd	G
12	Chunk	P	Corporation	G
13	Difference	R	Area	P
14	Hit	I	Measure	P
15	Mistake	I	Chunk	P
16	Issue	I	Piece	P
17	Idea	I	Part	P
18	Step	R	Scale	R
19	Trouble	I	Percentage	P
20	Reason	M	Proportion	P
21	Surprise	I	Portion	P
22	Event	I	Body	G
23	Part	P	City	S
24	Money	M	Window	M
25	Man	H	Onion	C
26	Guy	H	Pot	C
27	Fan	H	Egg	C
28	City	S	Bowl	C
29	House	S	Saucepan	C
30	Toe	H	Skillet	C

A=amount; C=culinary; G=group/collection; H=human; I=issue/situation; M=miscellaneous; P=part/portion/measure; S=space/place

Table 8 shows the numeric breakdown of each of the nouns into the semantic domains found in the data for both *big* and *large*, and Figure 3 offers a visual representation of this distribution.

Table 8. A numeric breakdown of the semantic domains into which each of the nouns has been classified (type frequency). Bold indicates a notable difference between the two adjectives.

	big	**large**
A	0 (0%)	5 (17%)
C	0 (0%)	6 (20%)
G	4 (13%)	6 (20%)
H	4 (13%)	0 (0%)
I	**13** (43%)	0 (0%)
M	3 (10%)	1 (3%)
P	2 (7%)	**9** (30%)

R	2 (7%)	2 (7%)
S	2 (7%)	1 (3%)

A=amount; C=culinary; G=group/collection; H=human; I=issue/situation; M=miscellaneous; P=part/portion/measure; S=space/place. Percentages are relative to the 30 most frequently collocating nouns with each adjective, given for quick numeric perspective; total percentage may be higher than 100 due to rounding errors.

Figure 3. Distribution of semantic domains for nouns collocating with *big* and *large* (type frequency).

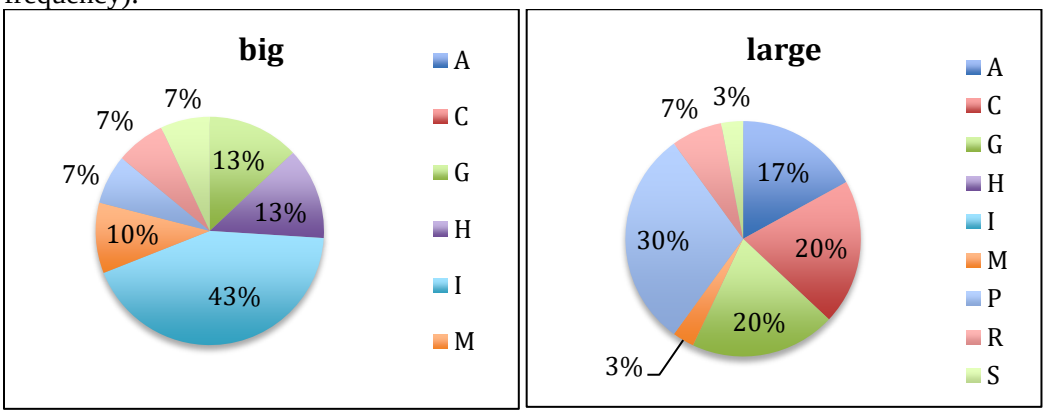


Table 8 and Figure 3 each show the distribution of the 30 most frequently collocating nouns with the adjectives *big* and *large* in terms of types. Analyzing the collocation data in this way—as with *huge* and *enormous*—allows us to point to specific words and their associated semantic categories that tend to collocate with either *big* or *large*, or in some cases both adjectives. Again, by examining token frequencies, we may gain insight into the strength of association between the target adjectives and the semantic categories of collocating nouns—and also of specific collocations. With this in view, Table 9 and Figure 4 display the distribution of the semantic domains for nouns that tend to collocate with *big* and *large* in terms of token frequency. Overall, the distribution of collocating nouns across the semantic categories is similar between type and token frequency, but a notable difference is seen in "amount" (A) nouns collocating with *large*. When looking at type frequency, 17% of the 30 most frequent nouns fall into this category, but when looking at token frequency the percentage increases to 33%.

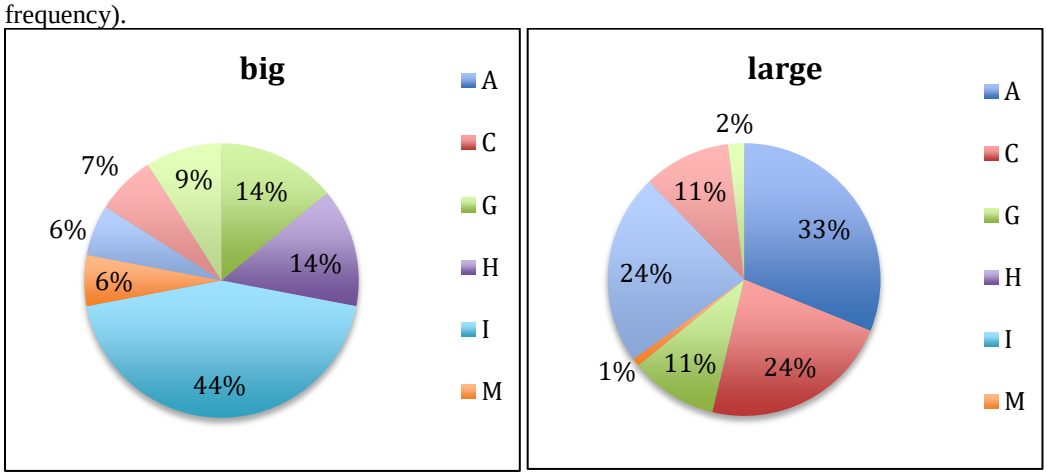
Addressing differences between type and token frequency is an important issue—one which is beyond the scope of this study but which should be addressed in future work.

Table 9. A numeric breakdown of the semantic domains into which each of the nouns has been classified (token frequency). Bold indicates a notable difference between the two adjectives.

	big	**large**
A	0 (0%)	**9616 (33%)**
C	0 (0%)	**7131 (24%)**
G	3544 (14%)	3295 (11%)
H	3632 (14%)	0 (0%)
I	**10986 (44%)**	0 (0%)
M	1624 (6%)	340 (1%)
P	1399 (6%)	**6868 (24%)**
R	1683 (7%)	3295 (11%)
S	2260 (9%)	623 (2%)

A=amount; C=culinary; G=group/collection; H=human; I=issue/situation; M=miscellaneous; P=part/portion/measure; S=space/place. Percentages are relative to the 30 most frequently collocating nouns with each adjective, given for quick numeric perspective; total percentage may be higher than 100 due to rounding errors.

Figure 4. Distribution of semantic domains for nouns collocating with *big* and *large* (token frequency).



Each of the four dictionaries broadly represents each of the adjectives as marked by great size or degree; some specify further by way of exemplification and others do not. Table 10 shows the lexicographic results for *big* and *large*. In the *CALD*, the semantic domain of "issue or situation" is accounted for in the entry for *big*. The fourth sense is

"important", and this is what is intended by *big* when it collocates with nouns of this kind. In fact, the entry gives the following examples of potential collocates: *decision*, *difference*, and *story*. This semantic category is not given in *CALD*'s entry for *large*. *COBUILD* offers a different representation, however. In addition providing the "important" sense for *big*, it is also provided this sense for *large*, which is not corroborated by the list of most frequent nouns that collocate with *large* in *COCA*. Perhaps, however, this sense is found in nouns that less frequently collocate with *large*. Future work should incorporate a more substantial list of nouns collocating with target adjectives. This could reveal the most important semantic categories that should be represented in a dictionary, and could also help to refine the presentation of definitions in connection with frequency patterns in current language use.

Table 10. Lexicographic results for collocational patterns.

		BIG	LARGE
CALD	SITUATION/ISSUE	√	
	PART/PORTION		
COBUILD	SITUATION/ISSUE	√	X
	PART/PORTION		
OALD	SITUATION/ISSUE	√	
	PART/PORTION		√
LDOCE	SITUATION/ISSUE	√	
	PART/PORTION		√

√= semantic category for nouns found in corpus data is corroborated in the dictionary consulted
X = semantic category for nouns found in corpus data is found in the dictionary consulted when it should not be; no mark indicates that there is no representation of the semantic category found in the dictionary consulted

Notably, *large* is not listed as a synonym for this sense of *big* on Thesaurus.com. The *LDOCE* offers entries similar to those in the *CALD*; however, it goes a step further to specify in the entry for *large* the following note: "Large implies broadness, generosity,

and comprehensiveness (large view). Big suggests importance and impressiveness but often in contrast to genuine worth". This is one of few places where explicit reference to the semantic domain with which a word is associated is found. Finally, the *OALD* lists "important" as the first sense for *big*, providing examples that can be categorized as "issue/situation" (e.g., "decision"), and does not offer it as a sense for *large*.

With respect to the other point of interest in the *big*/*large* data—the "part, portion, or measure" category, the account is simple. In none of the four dictionaries does the entry for *big* offer a sense pertaining to "part, portion, or measure". Given that there are some instances of such collocations, one could not argue with this sense being represented in the entries for *big*, but neither can one justifiably criticize its absence. While the representation of this sense is expected in the entries for *large*, this expectation is not quite met. This sense is not given in either the *CALD* or the *COBUILD*. It is given by way of exemplification in the *OALD*: "proportion" is listed as a potential collocate. The *LDOCE* goes a step further, however, not only listing "portion" as a potential collocate but also specifying that *big* is used for bulk or mass amounts and *large* for physical dimensions, quantities, amounts, or capacities. "Parts and portions", the category to which the collocating nouns here belong, belong to the physical domain. Furthermore, depending on how one categorizes, one could merge the amount category with the part/portion category; it also includes nouns that belong to the quantifiable physical domain. In addition, when one looks beyond these two categories, we find also the culinary category—the most specific of those included here, which consists of specifically physical items.

4. The nature of near-synonymy in adjectives

Analysis of the corpus data shows some—although not overwhelming—predictability in the semantic domains of the nouns with which the selected near-synonymous adjectives collocate. This result is expected given the adjectives under consideration are similar with nuanced differences, rendering them likely to lead to similar results with differences in the finer details. Where there is predictability, it is sometimes, but not always, reflected in the lexicographic entries examined for this study. Despite inconsistency in the representation of near-synonyms discovered in and across the dictionaries consulted, the lexicographers should not be faulted. Distinguishing near-synonyms is challenging because of their very nature of being similar yet different in very subtle ways. Let us consider how the nature of near-synonymy might contribute to slippery patterns in collocation.

The distinct senses of the near-synonymous adjectives were not considered in the initial collection of data, and this is important for understanding some of the collocational patterns that emerged. Because it is a particular sense of a given adjective—rather than the lexical unit—that is nearly synonymous with a particular sense of another given adjective, the semantic domain of nouns with which each adjective collocates is affected by which sense is intended. A quick search using Thesaurus.com brings up *considerable*, *powerful*, and *significant* as synonyms for *big* under the "important" sense, which is the second on listed. Under the first sense, the sense that corresponds to the adjectival senses compared in this study, other synonyms are listed: *enormous, huge, fat, colossal*, and so on. One might expect that the semantic domain of a noun collocating with any of these adjectives will be different depending on which sense of the adjective is modifying it, but

possibly it could also be argued that the sense of the adjective in use might be fed by the noun it modifies.

In Murphy's (2010) typology of adjectives, she describes gradable adjectives, the category under which those addressed here fall. Gradable adjectives can have meaning at different strengths and different values. As it applies here, "bigness" varies, and the adjectives here studied were paired upon this idea. That is, *big* and *large* were paired because they both have relatively neutral meanings in terms of their "size" sense; *enormous* and *huge* both indicate something yet bigger is being described. Murphy (2010: 222-23), furthermore, writes, "What adjectives can denote is notoriously fluid, since their interpretation is tied up with the interpretation of the nouns that they modify." Adjectives typically denote properties, and those properties belong to something. Because of this, the meaning of an adjective can shift in relationship with the noun it modifies. Murphy offers the examples *long snake* and *long DVD*. In the former, the adjective denotes a physical measurement; in the latter, the adjective refers to the duration of time. In most cases, dictionaries provide a different sense for each of these meanings, but when the difference in meaning between these two uses of *long* is examined, the way that meaning emerges differently from the adjective depending on the semantic detail of the noun it modifies is clear.

While distinct senses for *big* are listed in most dictionaries (e.g., larger than average size, important), the line to be drawn between these senses can be fuzzy. *Big*, as it is intended to convey importance, gains its meaning by extension from the sense indicating physical size. If we look at the collocation *big question*, in which the adjective bears its "important" sense, the same antonym (i.e., "small") that applies to the adjective

indicating physical size also applies in the sense of importance. Since, following Murphy (2010), an adjective's meaning can shift in relationship with the noun it modifies, one might wonder whether or not sense distinction matters in consideration of synonymous relations. However, since not all senses of a particular adjective are necessarily synonymous with those of another adjective, sense distinction must be important in the depiction of near-synonymous relations. At the same time, it seems clear that the semantic content of nouns modified by a particular adjective contributes to the sense of an adjective in use.

Indeed near-synonyms bear subtle semantic differences, and sense distinction provides only one perspective for analysis. Cruse (1986: 266) writes that there is "no neat way of classifying synonyms" but comes to the following conclusion: "synonyms must not only manifest a high degree of semantic overlap, they must also have a low degree of implicit contrastiveness". Elsewhere, he explains that near-synonyms are identical in terms of their "central" traits but differ in terms of their "peripheral". This is a useful place to begin, but a yet finer degree of nuance needs to be addressed. Edmonds and Hirst (2002) acknowledge Cruse's (1986) description of near-synonymy but argue for clearer refinement of how exactly near-synonyms differ, toward creating a computational model capable of handling near-synonymy. Specifically, the authors present the representation of lexical knowledge in terms of granularity, describing in detail an analysis of near-synonymy with the objective of creating an algorithm in a model of the lexicon.

At the heart of Edmonds and Hirst's (2002) analysis is, first, a distinction between specificity and granularity in meaning and, second, between fine- and coarse-grained aspects of denotation. With respect to the former, a thing can be specific yet either fine-

or coarse-grained, with grainedness as a property of representation in the lexicon, not of the concept itself. The distinction between fine- and coarse-grained aspects is somewhat analogous to Cruse's (1986) distinction between central and peripheral traits. Coarse- and fine-grainedness refers to the aspects or dimensions of a lexical item that, respectively, draw the semantic similarity to and differentiate it from another lexical item. These aspects can include denotation, style, expressivity, and structure (Cruse 1986).

Edmonds and Hirst (2002: 106) present a clustered model of lexical knowledge, in which a set of near-synonyms—like those under investigation here—are clustered together under "a common coarse-grained meaning", which "provides a mechanism for representing finer-grained aspects of denotation, attitude, style, and usage that differentiate the near-synonyms in a cluster". A cluster is comprised of overlapping categories between lexical representations. This is Edmonds and Hirst's (2002) response to questions about how to apply Cruse's (1986) dichotomy of central and peripheral traits; that is, which traits should be counted central and which peripheral? According to Edmonds and Hirst (2002: 116-117), "[n]ear-synonyms can occur at any level of specificity, but crucially it is the fine granularity of the representations of their meanings that enables one to distinguish one near-synonym from another." As a consequence, in order for a conception of near-synonymy to be sufficient, it must incorporate the notion of granularity. This is an important refinement of how near-synonyms should be treated in analysis. With the introduction of granularity of representation, subtle distinctions between the meanings of near-synonymous words can be accurately defined and accounted for, as can interrelated fine- and coarse-grained aspects of meaning between a set of near-synonyms.

This idea is relevant to the current study because it explains and visually depicts the nature of the subtle complexity of near-synonymy as a lexical relation and points to a clear way in which near-synonymy affects the structure of the lexicon. Near-synonyms are complex because they are similar in terms of shared coarse-grained aspects but differ in terms of fine-grained aspects, and homing in on the subtle fine-grained representations that distinguish near-synonymous words can be challenging. The issue is increasingly complicated by "several interrelated near-synonyms with interrelated differences" (Edmonds & Hirst, 2002: 117). This is precisely the case of the near-synonyms being examined in this study, which helps to explain why the trends found among the corpus data are not clear-cut.

This idea of granularity in representation, examined in view of Murphy's description of the fluid properties of an adjective, leads to a clearer idea of how to approach the subtle differences in meaning between near-synonymous adjectives. It additionally reveals why this refinement is important and challenging in terms of both the structure of the lexicon and lexicographic representation. Not only are the fine-grained aspects of near-synonymous adjectives significant for discerning subtle nuances in meaning—both shared meaning and points of departure; so are the collocational patterns because nouns modified by an adjective contribute to the meaning of their modifiers.

5. Discussion and conclusions

5.1 Lexicographic representation of near-synonyms

In general, *big* and *large* are somewhat distinguishable in learner's dictionaries. The *OALD* and *LDOCE* most clearly distinguish these adjectives in accordance with the

corpus data. The *CALD* does not reflect the semantic nuances found for *large*, and the *COBUILD* commits an error in its entry for *large*, in view of the corpus data.

Looking to *huge* and *enormous*, the only dictionary that accurately captures the semantic nuance in these adjectives is the *COBUILD*, which is in contrast with what was found for *big* and *large*. The *OALD* is on the right track for *huge*—but only for *huge*, and the *LDOCE* incorrectly represents the semantic nuance found for *huge*. There is otherwise nothing to report for either of the adjectives. It is not easy to distinguish between these adjectives in the dictionaries used in this study. Further investigation of additional semantic traits will help to be able to conclude more soundly whether the dictionaries consulted represent these two adjectives in a distinguishable way.

Overall, little consistency was found within a given dictionary with respect to accurate representation of corpus patterns, for either of the two pairs of adjectives studied, and neither is there much consistency across the four dictionaries consulted. One of two conclusions can be drawn. The first is that more accurate and systematic methodology attending more closely to corpus patterns in the construction of learner's dictionaries is required. The second is that the methodology of the present study—and others like it—need to be refined in order to reap more accurate results.

There are some limitations to this study, which should be kept in view for future work. It could be useful to look at a greater number of nouns collocating with the defined set of near-synonymous adjectives in order to increase the amount of data and provide a more complete analysis, particularly with respect to capturing the full scope of possible meanings and senses that should be reflected in a dictionary entry. Furthermore, examining a broader collocational range—allowing for collocations beyond only

immediately collocating adjectives and nouns—will also be informative. Finally, a more systematic breakdown of the senses of each of the near-synonymous adjectives would allow for finer analysis along this parameter, which could cast a different light on the issue of predictability of the use of a particular near-synonym in a particular environment and the semantic domains of typically collocating nouns. While a corpus like *COCA* is valuable for a study of this nature, one thing it cannot do is distinguish between senses of a given adjective, so this must be done manually. Achieving a systematic way of doing this will greatly enhance future work.

5.2 The nature of near-synonymy

Near-synonymous adjectives can be generally described by the epigraph from Gerard Manley Hopkins. They are near enough, within the lexicon, to have something in common but not near enough to be the same. The analysis here allows us to begin to conclude that, in fact, the structure of the lexicon is affected by synonymous relationships between words. If we envision a cluster in which the lexical unit *big* has multiple senses, each of which connect to some of the same but some different near-synonyms, then aspects of the lexical item *big* are structurally connected to aspects of other lexical items, but not the same aspects for each of the other lexical items. Furthermore, if the noun that an adjective modifies contributes meaning to an adjective, then we also have to conclude that the adjective is not contained in static representation within the lexicon but rather can shift as its semantic properties shift in usage. This offers support for usage-based models of language (Barlow & Kemmer, 2000; Boyland, 2009; Langacker, 1987), an angle not pursued here but which warrants focus.

If an adjective's meaning is fluid and stabilizes its meaning by way of the noun it modifies, as Murphy (2010) claims, the properties of the nouns that contribute to an adjective's semantic nature must be determined, and frequency effects should be incorporated. A collocation study, in conjunction with one of a lexicographic nature, can begin to do this effectively. What must be kept in view is the constant process of subtle changes in usage, which inevitably play at the fuzzy boundaries of meaning in near-synonymous relationships between adjectives.

In addition to refining the methodology, as described in the previous subsection, future research should address additional sets of near-synonymous adjectives in order to compare and contrast how the adjectives behave. This would contribute to an understanding of near-synonymous adjectives in general and also models of the lexicon aimed at handling this issue. Furthermore, future work should incorporate ideas from usage-based models of language and the broader field of cognitive science to gain insight from research on concept and category organization. Abundant research in cognitive science runs analogous to the clustered model of representation, namely with respect to categories and overlapping features between category members, which can enhance our understanding of the structure of the lexicon and the place of near-synonyms therein.

References

Barlow, M., & Kemmer, S. (2000). *Usage-based models of language*. Stanford, CA: CSLI Publications.

Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied linguistics*, *15*(2), 169-189.

Boyland, J. T. (2009). Usage-based models of language. In D. Eddington (Ed.), *Experimental and quantitative linguistics*, 351-419. Munich: Lincom.

Cambridge advanced learner's dictionary (3rd ed.). (2008). Cambridge, UK; New York: Cambridge University Press.

Collins COBUILD English grammar (2nd ed.). (2005). Glasgow, Scotland: HarperCollins.

Cruse, D. A. (1986). *Lexical semantics*. New York, NY: Cambridge University Press.

Davies, M. (2008-) *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

Divjak, D. (2010). *Structuring the Lexicon: A Clustered Model for Near-synonymy*. Berlin: De Gruyter Mouton.

Divjak, D., & Gries, S. T. (2008). Clusters in the mind?: Converging evidence from near synonymy in Russian. *The Mental Lexicon*, *3*(2), 188-213.

Edmonds, P. (1999). Semantic Representations of Near-Synonyms for Automatic Lexical Choice. Ph.D. Dissertation, Department of Computer Science, University of Toronto.

Edmonds, P. & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics, 28*(2): 105-144.

Evans, V. and Green, M. (2006). Cognitive linguistics: An introduction. :PUBLISHER

Firth, J. (1968). "A synopsis of linguistic theory 1930-1955". In F. R. Palmer (ed.): Selected Papers of J. R. Firth 1952-59 ,(1-32). Bloomington, IN: Indiana University Press.

Gries, S. Th. (2010). Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, *5*(3), 323-346.

Gries, S. T., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, *34*, 121-150.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.

Langacker, R. (1987). *Foundations of cognitive grammar, Vol. I: Theoretical prerequisites.* Stanford, CA: Stanford University Press.

Laufer, B. (1990). Ease and difficulty in vocabulary learning: Some teaching implications. *Foreign Language Annals, 23*(2), 147-155.

Lipka, L. (1990). *An outline of English lexicology: lexical structure, word semantics, and word-formation* (Vol. 3). Walter de Gruyter.

Longman dictionary of contemporary English. (1984). Harlow, England: Merriam-Webster.

Martin, M. (1984). Advanced Vocabulary Teaching: The Problem of Synonyms*. *The Modern Language Journal*, *68*(2), 130-137.

Murphy, M. L. (2010). *Lexical meaning*. Cambridge University Press.

Oxford advanced learner's dictionary of current English (7th ed.). (2005). Oxford; New York: Oxford University Press.

Princeton University 2010. About WordNet. *WordNet*. Princeton University. Available online at http://wordnet.princeton.edu/.

Roget's international thesaurus (5th ed.). (1992). New York: HarperCollins.

Roget's 21st century thesaurus (3rd ed.). Retrieved December 09, 2014, from Thesaurus.com website: http://www.thesaurus.com/browse/.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers.

Taylor, J. R. (2002). Near-synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Sciences, 25*, 263-284.

Verlinde, S., & Selva, T. (2009). Corpus-based versus intuition-based lexicography: defining a word list for a French learners' dictionary. *coach, 83*, 1424.

Xiao, R., and McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied linguistics, 27*(1), 103-129.

Appendix

A. Lexicographic information for *big* and *large*. The entry for each is adapted from each respective dictionary. Collocating nouns extracted from examples are given in parentheses following the definitions. Synonyms are presented in brackets following the parentheses.

	BIG	**LARGE**
CALD	1. Large in size or amount. (man, shoes, house, eyes, pay rise, slice, race) [large] 2. To a large degree. (spender, eater) [large] 3. Used to add emphasis. (bully) informal [large] 4. Important, because of being powerful, or having a lot of influence or a serious effect. (decision, difference, story, banks) [important]	1. Big in size or amount. (house, computer manufacturer, car, number, fall (drop), survey, emergency food-aid)
COBUILD	(5-diamond frequency) 1. A big person or thing is large in physical size. (country, husband, car) [large] 2. Something that is big consists of many people or things. (contingent, backlog) [large] 3. If you describe something such as a problem, increase, or change as a big one, you mean it is great in degree, extent, or importance. (problem, increase) [serious] 4. A big organization employs many people and has many customers. (banks, companies) [large]	(5-diamond frequency) 1. A large thing or person is greater in size than usual or average. (rivers, lakes, room, man) [big] 2. A large amount or number of people or things is more than the average amount or number. (amount, number, figures) [not small] 3. A large organization or business does a lot of work or commercial activity and employs a lot of people. (company, organization) [big] 4. Large is used to indicate that a problem or issue which is being discussed is very important or serious. (problem, question) [serious]
LDOCE	1. Of great force. (storm) 2a. Large in dimensions, bulk, or extent; also large in number or amount. (house, fleet) 2b. Conducted on a large scale. (business) 2c. Important in influence, standing, or wealth. 6a. Chief, outstanding. (issue) 6b. Outstandingly worthy or able. (person) 6c. Of great importance or significance (moment)	1. Having more than usual power, capacity, or scope; comprehensive. (social justice) 2. Exceeding most other things of like kind, esp in quantity or size; big 3. Dealing in great numbers or quantities; operating on an extensive scale (business, exporter) [large, big, and great are often interchangeable; large is often preferred to describe physical dimensions, quantity, amount, or capacity (garden/portion/number/ bucket). Big is more appropriate for

		bulk, mass, or weight and for volume of sound. Used figuratively, big, large, and great have similar connotations to their literal use. Large implies broadness, generosity, and comprehensiveness (large view). Big suggests importance and impressiveness but often in contrast to genuine worth. Compare HUGE].
OALD	1. Large in size, degree, amount, etc. (man, house, increase, shirt, computer company, grin, money, blow) [large] 2. Important, serious (only before noun) (decision, match, mistake, moment, boys) [important] needing a lot of effort, money, or time to succeed (ideas) [ambitious] 3. Popular with the public; successful (color, band) [popular] enthusiastic about someone or something (fan) [enthusiastic] 4. Doing something often or to a large degree (eater, drinker, spender) [doing something a lot]	1. Big in size or quantity (area, family, house, car, appetite, number, sums, child, proportion, share, producer, lady (=fat)) Clothing size 2. Wide in range and involving many things (issue, scale, picture)

B. Lexicographic information for *enormous* and *huge*. The entry for each is adapted from each respective dictionary. Collocating nouns extracted from examples are given in parentheses following the definitions. Synonyms are given in brackets following the parentheses.

	ENORMOUS	**HUGE**
CALD	1. Extremely large. (car, house, salary, I, help)	1. Extremely large in size or amount. (house, costs, number, successes)
COBUILD	(3-diamond frequency) 1. Something that is enormous is extremely large in size or amount. (bedroom, amount) 2. You can use enormous to emphasize the great degree or extent of something. (disappointment, capacity)	(4-diamond frequency) 1. Something or someone that is huge is extremely large in size. (glasses, piece of canvas, driver) [gigantic] 2. Something that is huge is extremely large in amount or degree. (number, profits) [enormous] 3. Something that is huge exists or happens on a very large scale, and involves a lot of different people or things. (problem, scale) [vast]
LDOCE	1. Marked by extraordinarily great size number, or degree; esp exceeding usual bounds or expectations 2a. Exceedingly wicked; shocking (sin) 2b. *archaic* abnormal, inordinate [huge is synonymous]	1. Great in size, scale, degree, or scope; enormous. (mountains, success) [huge, vast, immense, enormous, mammoth, elephantine, giant, gigantic, colossal, gargantuan, titanic: huge is a general term, expressing great size, bulk, or capacity (man, pile). 2. Enormous suggests size or degree far in excess of what is usual]
OALD	1. Extremely large. House, dog, amount of time, interest, problems [syn: huge, immense]	1. Extremely large in size or amount; great in degree (syn: enormous, vast) (crowd, eyes, debts, amount of data, sums of money, problem) 2. Very successful (informal) (band)