



universität
wien

MASTERARBEIT

Titel der MASTERARBEIT

“Synonymy, plesionymy and sameness of meaning:
A corpus-based behavioral profile analysis of adjectives
of SIZE”

Verfasser

Christian Moser, BA

angestrebter akademischer Grad

Master of Arts (MA)

Wien, im April 2014

Studienkennzahl lt. Studienblatt: A 066 812

Studienrichtung lt. Studienblatt: English Language and Linguistics

Betreuer: Univ.-Prof. Dr. Malgorzata Fabiszak

Declaration of authenticity

I confirm to have conceived and written this Master's thesis in English all by myself. Quotations from other authors and sources, as well as any ideas borrowed and/or passages paraphrased from the words of other authors are all clearly marked within the text and acknowledged in the bibliographical references.

Vienna, April 2014. Signature: _____

**This thesis is dedicated to my grandfather
In loving memory, I wish you were still here**

Acknowledgements

First and foremost, I would like to express my gratitude to Dr. Malgorzata Fabiszak who not only taught me a thing or two about linguistics throughout the course of my studies, but was also willing to supervise my thesis, guide me through the various stages of composing a scientific paper and generally was on hand with help and advice. Thank you, Professor Fabiszak, I really appreciate it.

Furthermore, I want to thank my parents, Johanna and Hans Moser. They are very special people and if it was not for them, their support, their love and kindness, I would neither be the character I am today nor would I be where I am today. They are my greatest supporters and harshest critics at the same time and on top of that, they are my friends. Thank you, mom and dad, for your continuing encouragement and your faith in me. I owe you, eternally.

I am deeply indebted to Verena, the love of my life. No one has been more solidly in my corner than her. As cliché as it may sound, she always is, was and has been there for me in prosperity and adversity. I am a proud and complex man and therefore it is not always easy to bear with me but she always indulged my every whim. She makes me feel home wherever we are and she substantially contributed to molding the person I am today and I thank her for that. V, you rock my world.

My lacrosse team. You guys are complete buffoons and I love you for that. Much like absolute synonymy, friends like you are beyond rare. I am grateful to have met you and want to say thank you for our sincere friendship and all the fun we have had. Here is to the future and the best that is yet to come. You are the regulars of my favorite years. Go Monarchs! Head, Heart, Hustle.

My friends back home in Carinthia. Sometimes I might have my head up high in the clouds but they are always there to put my feet back on the ground and remind me that you can take the boy out of the country but you cannot take the country out of the boy. Thanks.

My friends near and far, stateside and elsewhere. You enrich my life and without you it would not be half as fun. God bless you.

Last but not least, thanks to the rest of my family in Carinthia and Styria and also to my step-family in Upper Austria. Thanks for treating me like one of your own and having accepted me as a member of the family. Walter, thank you for your generosity, I cannot emphasize strongly enough how grateful I am. Manuel and Michael, I appreciate your friendship and cannot wait for our next lukewarm beer. Cheers.

Abstract

This thesis investigates the synonymy of three frequently used adjectives belonging to the semantic field of SIZE, *big*, *large* and *huge*, three lexical items considered to be synonymous by various thesauruses and dictionaries. Its aim is to provide an account of the synonymy of the three adjectives from both, a semantic and syntactic point of view by utilizing a behavioral profile analysis.

In order to do so, raw frequency information on the adjectives has been gathered from the Corpus of Contemporary American English, close concordance searches and statistical collocation analyses have been carried out. A significant number of tokens have been manually annotated and analyzed. The language R statistical environment has been incorporated to conduct statistical tests, process the data and graphically present the findings. Finally, the findings have been compared to and interpreted with regard to previous studies.

The study shows that even the adjectives *big* and *large*, two semantic entities canonically thought to be as synonymous as possible, are, in fact, propositional synonyms, near-synonyms or plesionyms at best. However, they can be considered more synonymous in terms of syntactic properties than concerning semantic features, for instance, and can therefore be considered as actual synonyms in certain respects. Moreover, while sharing a number of attributes with *big* and *large*, *huge* is, linguistically speaking, distinctly different from the other two adjectives. The study reinforces the notion that cognitive synonymy, sameness of meaning in any and every syntactic, semantic and pragmatic aspect is merely impossible but proposes that partial synonymy actually exists, especially within certain paradigms such as generative linguistics or autonomous linguistics. Synonymy is not a clear-cut linguistic category and synonyms, near-synonyms and plesionyms form a continuum full of fine nuances of meaning rather than being finite and well-defined entities.

Keywords: cognitive linguistics, semantics, synonymy, corpus-based, adjective

Table of contents

Abstract	IX
1 Synonymy	1
1.1 What is synonymy?	2
1.2 Definitions of synonymy	3
1.2.1 Synonymy in lexicography	3
1.2.2 Synonymy in linguistics.....	7
1.3 Motivation	14
1.4 Theoretical framework and previous studies	17
1.5 Aim of the paper and hypothesis.....	24
2 The study	26
2.1 Source of data and choice of method	26
2.1.1 The Corpus of Contemporary American English	28
2.1.2 Behavioral profile analysis	29
2.1.3 Dictionaries and thesauruses	31
2.2 Supporting Software.....	32
2.2.1 Language R, statistical computing.....	32
2.3 Data selection, filtering and annotation	33
2.4 Choice of statistical tests and graphical representation	37
3 Results	40
3.1 Lexicographical analysis	40
3.2 Behavioral profile analysis.....	46
4 Discussion and theoretical implications.....	66
5 References.....	68
6 Appendices	72
7 Abstract in German	87

List of tables

Table 1: Dimensions of synonymy.....	12
Table 2: Raw frequencies of <i>big</i> , <i>large</i> and <i>great</i> in a 5.7-million-word sample from the Longman-Lancaster corpus	19
Table 3: ID-Tags for the behavioral profiles of <i>big</i> , <i>large</i> and <i>huge</i>	34
Table 4: Excerpt co-occurrence table <i>big</i> , <i>large</i> and <i>huge</i>	36
Table 5: Lexicographical sources examined.....	72
Table 6: Raw frequencies of <i>big</i> , <i>large</i> and <i>huge</i> in the Corpus of Contemporary American English	46
Table 7: Five most frequent collocates of <i>big</i>	50
Table 8: Five most frequent collocates of <i>large</i>	50
Table 9: Five most frequent collocates of <i>huge</i>	51
Table 10: Excerpt of the ID-Tags for <i>big</i> , <i>large</i> and <i>huge</i>	54
Table 11: Excerpt of the R dataframe	55
Table 12: Excerpt of the R pivot table.....	56

List of figures

Figure 1: Snakeplot of differences between BP vector values <i>big</i> - <i>large</i>	23
Figure 2: Multiple correspondence analysis full dataset	58
Figure 3: Multiple correspondence Analysis BP <i>big</i>	59
Figure 4: Multiple correspondence Analysis <i>large</i>	60
Figure 5: Multiple correspondence Analysis <i>huge</i>	62
Figure 6: Cluster Analysis <i>big</i> , <i>large</i> and <i>huge</i>	64
Figure 7: Syntactic Cluster Analysis <i>big</i> , <i>large</i> and <i>huge</i>	65

1 Synonymy

Synonymy is arguably one of if not the most investigated lexical relation in linguistics. A seemingly endless number of books, theses and articles are dedicated to the very topic and examine the phenomenon from almost any and every angle imaginable. Yet this specific linguistic area has hardly lost any of its appeal to linguists and is still subject to numerous investigations. Considering synonymy's popularity in linguistics and the vast number of studies conducted, one may ask what the point of yet another thesis on synonymy and sameness of meaning would be. The answer to that question is very simple and straightforward: not everything has been said. In fact, there is still plenty of room for improvement of the category synonymy and all its subcategories based on actual evidence. The advent of new methodologies such as the Behavioral Profile Analysis, rendered possible by scientific progress associated with the introduction of electronic corpora, allow for unprecedented, multifactorial, fine-grained and all-encompassing analyses of synonymy previously unheard of. The paper at hand utilizes data gathered from a state-of-the-art corpus to, first and foremost, establish the behavioral profiles of the lexical items *big* and *large*, canonically thought to be synonymous. Additionally the profile of another very similar lexeme, *huge*, is examined. Subsequently the findings are interpreted, compared to previous studies, similarities and differences are highlighted and explained. In essence, this study seeks to enhance previous accounts of synonymy and its subcategories by investigating the lexical items *big*, *large* and *huge* and categorizing them within the linguistic framework of synonymy.

The following introductory sections establish the linguistic notion of synonymy, define and discuss synonymy in both lexicography and linguistics and determine the motivation of the study at hand. Chapter 2, the empirical part of the study, presents the findings as well as their theoretical implications and discusses the limitations of the study at hand as well as limitations of behavioral profile analyses and corpus-based approaches to lexical relations.

1.1 What is synonymy?

“Synonymy is formal variation of a concept-function” (Glynn 2012: 208).

One aspect of synonymy is that it is a particular type of classification, a paradigmatic relation between two or more semantic items or concepts. To paraphrase, synonymy is a certain type of linguistic categorization and, as such, has received a lot of attention from linguists and laypeople alike. George Lakoff, one of the founding fathers of cognitive linguistics, states that:

Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action and speech. Every time we see something as a kind of thing, for instance, a tree, we are categorizing. [...] Without the ability to categorize, we could not function at all, either in the physical world or in our social and intellectual lives. (1987: 5-6)

Murphy (2003: 134) notes that synonymy is usually defined by determining what linguistic properties certain words share and to what extent as opposed to how they differ which would be the area of interest for linguistic investigations on antonymy. In another paper, Murphy (2013: 279) suggests that while synonymy is a metalexic relation, which means it is knowledge about the language rather than knowledge of the language, speakers of English are comfortable with not only the term but also its meaning. Furthermore she claims that synonymy is one of very few metalinguistic terms used by laypeople in everyday language.

These two reasons alone, the importance of categorization along with the common use of the term may already provide sufficient grounds for the almost gapless linguistic coverage of the issue. However, Gries and Otani (2010: 124-125) propose yet another possible reason for the popularity of synonymy among linguists. Deese's (1964) investigation on the associative structure of common English adjectives, one of the first empirical studies on synonymy, concluded that *big* and *large* might be two of the most synonymous concepts in the English language. Deese's study did not only yield countless further investigations on

synonymy and, consequently, antonymy, and sparked numerous methodologies and types of inquiry, it also established the lexical field of SIZE as the default for studies on synonymy. Needless to say, even though over the years linguists have cast their nets out wide and advanced into other, different semantic fields and word classes using ground breaking methodologies, Deese's empirical work is still considered to be a standard reference.

As can be seen, the study of synonymy proves to be a highly fruitful area of linguistics in general and, since the advent of tagged electronic corpora, corpus linguistics in particular. I wholeheartedly agree with Gries and Otani (2010: 121) on the fact that synonymy is "the most frequently corpus-linguistically studied lexical relation". Nevertheless, despite being arguably the most investigated upon domain of linguistics, there are still aspects which have only been analyzed superficially if at all. More on what current research on synonymy within the lexical field of SIZE is lacking, which partly constitutes the motivation of this thesis, is to be found in section 1.3, motivation.

1.2 Definitions of synonymy

The general concept of synonymy has always been discussed controversially and different fields offer different definitions. In what follows, the representation of synonymy in lexicography and cognitive linguistics is presented and discussed.

1.2.1 Synonymy in lexicography

When it comes to the question of what it means to know a word there are several crucial elements that could be named. Spelling, pronunciation, implications, connotation, denotation, usage and so forth, but the most important property is very likely the meaning of the word. A good starting point for the search of word meanings are dictionaries and thesauruses. While these words are often used interchangeably, or synonymously, especially by dictionary users, Murphy (2013: 279) draws a quite clear distinction between orthographically organized

dictionaries and semantically organized thesauruses. Nevertheless she admits that with electronic access to dictionaries and thesauruses said distinction starts to fade and matters less since typically all a user has to do is to enter a lexical item into a search box.

Lexicographers and general, monolingual dictionaries define synonymy relatively straightforwardly as two or more lexical units with a shared meaning. Linguistic dictionaries usually add a bit of metalinguistic information. The definition provided by the Collins English Dictionary (CED) is presented in (1), (2) lists the Oxford Online Dictionary (OOL) entry. As for linguistic dictionaries, the information given by the Concise Oxford Dictionary of Linguistics (2nd ed.), (CODL) is depicted in (3) and the Oxford English Dictionary (OED) entry is listed in (4). All these dictionaries are available online and the entries were being retrieved online on January 14, 2014 via the University of Vienna. A close examination of the lexical resources is provided in table 5 in the appendix.

- (1) a word that means the same or nearly the same as another word, such as *bucket* and *pail*. (CED)
- (2) A word or phrase that means exactly or nearly the same as another word or phrase in the same language, for example *shut* is a synonym of *close*. (OOD)
- (3) The relation between two lexical units with a shared meaning. ‘Absolute’ synonyms, if they exist, have meanings identical in all respects and in all contexts. ‘Partial’ synonyms have meanings identical in some contexts, or identical only e.g. in that replacing one with the other does not change the truth conditions of a sentence. Thus paper is a partial though not an absolute synonym of article: compare I got my paper published, I got my article published. (CODL)
- (4) Strictly, a word having the same sense as another (in the same language); but more usually, either or any of two or more words (in the same language) having the same general sense, but possessing each of them meanings which are not shared by the other or others, or having different shades of meaning or implications appropriate to different contexts: e.g. *serpent*, *snake*; *ship*, *vessel*; *compassion*, *fellow-feeling*, *sympathy*; *enormous*, *excessive*, *immense*; *glad*, *happy*, *joyful*, *joyous*; *to kill*, *slay*, *slaughter*, *to grieve*, *mourn*, *lament*, *sorrow*. (OED)

N.B.: The Oxford Online Dictionary entry for synonymy actually reads “The state of being synonymous. (see synonym)” and the CED’s definition says “The character of being synonymous”. Therefore, in both cases, the entry presented above is, as a matter of fact, the entry for synonym. The lack of information on synonymy and the relatively restricted definition of the word synonymy within the CED and the Oxford Online Dictionary come as no surprise due to the fact that both these dictionaries are tailored towards the needs of laypeople and ordinary language users as opposed to experts and linguists. The OED and the Concise Oxford Dictionary of Linguistics, in turn, provide these kinds of metalinguistic information as they are intended for experts in the field.

An in-depth analysis of the four dictionary entries provides some major points for discussion which is yet another motive for the paper and the study at hand.

First of all, only the Oxford Online Dictionary and the OED mention that a necessary condition for two words in order to be considered synonymous is that they belong to the same language. The notion itself is not astonishing as, for two or more words to be absolutely synonymous, an important requirement is mutual interchangeability in any and every context which, when dealing with two different, not mutually comprehensible languages, obviously is not the case. However, it is remarkable that the non-linguistic Oxford Online Dictionary mentions the language-restrictedness of synonyms while an according passage is nowhere to be found in the Concise Oxford Dictionary of Linguistics. Murphy (2013: 282) states that “language-blind use of synonym is not unheard-of in linguistics”. In fact, she herself, in an earlier publication entitled “Semantic Relations and the Lexicon. Antonymy, Synonymy and Other Paradigms” (2003: 141), reviews ten definitions of synonymy and none of them considers the same language to be a necessary feature for synonymy. Nonetheless, even in cases where sameness of language is not explicitly mentioned, it is implied, Murphy concludes. Her taxonomy of synonymy and why, according to her, contrast is the general category for semantic relations will, among other paradigms and frameworks, be discussed in section 1.2.2 of the paper, which deals with sameness of meaning in linguistics. To sum up, while the word synonymy is de facto sometimes commonly used to classify two or more lexical realizations of

one concept in two different languages as pointed out by Murphy (2013: 282) and given in (5), from a linguistic point of view, for semantic entities in order to be considered synonymous, sameness of language is a crucial element. According to Murphy, such cases could be interpreted as translational equivalents rather than synonymous concepts.

- (5) The word *ghanika*, or food, is a synonym for 'yam.' (COCA: M. Lepowsky, 'The way of the ancestors', *Ethology* 30.3. 1991)

Secondly, the semantic qualities and requirements of synonyms mentioned in the dictionary entries under examination have to be considered. All four dictionaries, although differently phrased, convey the notion that synonymy is the relation between two or more linguistic items, that mean the same or nearly the same thing. Among linguists it is agreed upon the fact that absolute synonymy, sameness of meaning in any and every aspect, is exceptionally rare if at all possible and that, consequently, there have to be other types of close linguistic relationships similar to but not exactly the same as absolute synonymy. (cf. i.e. Cruse, 1986, 2002, 2004; DiMarco, Hirst and Stede, 1993; Murphy, 2003; Storjohann 2009) Bergenholtz and Gouws (2012: 309) note that among non-linguists the meaning of the term synonymy is usually limited to referring to a set of words which possess exactly the same meaning. Moreover, Bergenholtz and Gouws continue, general monolingual dictionaries, the type of dictionary customarily consulted by non-linguists, present information in a way that supports the aforementioned assumption of synonymy being defined as absolute sameness of meaning. While it may be true that some monolingual dictionaries define synonymy as absolute sameness of meaning or, at least, do not explicitly suggest different degrees of synonymy such as, for example, the Collins COBUILD British English Advanced Learner's Dictionary, most dictionaries, as a matter of fact, include a certain element of gradedness as evidenced by (1),(2),(3) and (4). The Collins COBUILD British English Advanced Learner's Dictionary's (COBUILD) entry, presented in (6), is of particular interest as it, despite being part of a corpus-based dictionary, lacks any hint towards gradedness whatsoever.

- (6) A synonym is a word or expression which means the same as another word or expression. (COBUILD)

It is worth mentioning that especially the use of corpora and corpus-based studies have considerably contributed to the most up to date definitions of synonymy and its subcategories, yet of all dictionaries a corpus-based one does not include gradedness of synonyms. A possible explanation could be that it is a dictionary aimed towards advanced learners and therefore only contains the most important information especially as the CED, which is also based on the COBUILD corpus, does include propositional and near-synonymy. Nevertheless, such a simplified presentation can be considered a striking, distinguishing feature.

The two linguistic dictionaries, the OED and the Concise Oxford Dictionary of Linguistics include, owed to their scientific character, additional, metalinguistic information on synonymy based on linguistic research which is not present in the CED and the Oxford Online Dictionary. In doing so, these dictionaries already touch upon a more elaborate, linguistic definition of synonymy.

In conclusion it can be said that all four dictionaries under investigation incorporate the current state of research in their definitions of synonymy in so far as they distinguish between the concept of actual synonymy and near-synonymy, meaning that two or more concepts can be used interchangeably within certain contexts but not ubiquitously. A comprehensive account of how synonymy and its sub-categories are currently treated within linguistics is provided in the following section.

1.2.2 Synonymy in linguistics

As mentioned above, it is widely accepted among linguists that absolute synonymy is “rare at best” (DiMarco, Hirst and Stede, 1993: 120) within the framework of cognitive linguistics due to the fact that “a word can express a myriad of implication, connotation and attitudes in addition to its basic ‘dictionary’ meaning.” (Edmonds and Hirst, 2002: 105). They then go on to assert that even if

absolute synonymy was theoretically possible, empirical and pragmatic data suggests that it would be exceptionally rare. Cruse further substantiates this notion by stating that “natural languages abhor absolute synonymy just as nature abhors vacuum” (1986: 270). Clark (1992) states that absolute synonyms within natural languages either develop new implications and nuances of meaning or simply cease to exist as there is no need for absolute synonyms whatsoever. Unfortunately natural languages do not work that way, at least not as distinct and clear cut as proposed and the semiotic ideal of one form only representing one meaning and one meaning only is not as straightforward in practice as it may seem in theory (Taylor, 2003: 263).

DiMarco et al. (1993) argue in a similar way. They suggest that while absolute synonymy, if at all possible, is yet to be proven and substantiated, sets or pairs of words that share a common denotation but differ in other terms of usage such as connotation or collocational constraints are not quite as scarce (1993: 120). Idiolectal language use and interpersonal variation are also factors to be considered when talking about synonymy and sameness of meaning. However, these aspects are beyond the scope of this mostly quantitative analysis of the synonymy of the words *big*, *large* and *huge* for a number of reasons, one of them being the fact that qualitative approaches are better suited to thoroughly investigate upon individual language variation.

With absolute synonymy, sameness of meaning in any and every linguistic aspect from denotation and connotation to syntactic environment, semantic relations and collocational constraints, constituting one end of a continuous scale of interchangeability and no shared properties whatsoever representing the other, there is plenty of room for other categories along the scale. Many linguists have proposed distinct frameworks and paradigms. The most important and experimentally reinforced ones, corpus-based or otherwise, are presented, explained and analyzed in the remainder of this chapter.

Two of the first linguists to elaborate on synonymy and to introduce and categorize different degrees of synonymy are Apresjan (1973) and Cruse (1986). In his book on lexical semantics (1986), Cruse elaborates on synonymy and

argues that, due to the fact that there is no obvious motivation for absolute synonyms in a language to exist, that the falsifiability of absolute synonymy, even among words canonically thought to be synonymous, is relatively straightforward. He supports his assumptions through quantitative analyses of usage differences of word pairs such as *begin* and *commence* or *hate* and *loathe*. Moreover, he identifies and introduces two central sub-categories of the lexical category that is synonymy, namely propositional synonymy and plesionymy. Cruse argues that pairs of absolute synonyms have to share all semantic and syntactic properties with each other. However, considering that absolute synonyms are beyond rare, consequently most if not all pairs of words thought to be synonymous have to differ in at least one respect. Cruse calls this special sub-category of synonymy propositional synonymy (1986: 270-280). Both, propositional synonyms and plesionyms, often called near-synonyms (Apresjan, 1973) distinguish themselves from the category of absolute synonymy in that they claim synonymy in every aspect imaginable within the paradigms of cognitive linguistics. The difference between propositional synonyms and plesionyms is a rather complex yet crucial one. Propositional synonyms are expressions which “can be substituted in any expression with truth-conditional properties without effect on those properties” (Cruse 2004: 155). Put another way, in expressions with truth-conditional properties such as statements that create a certain semantic framework, a word can be substituted without altering the truth conditions of said utterance and within this restricted environment considered to be absolutely synonymous. Plesionyms on the other hand, while still being very similar concepts, do, in fact, change the truth value of a statement. According to Storjohann, “[p]lesionyms are lexical items or constructions that designate very similar concepts and at the same time exhibit slight meaning differences so that they cannot be considered identical in meaning” (2009: 2140). DiMarco, Hirst and Stede’s (1993: 121) approach to near-synonymy and plesionymy is, although it is worded differently and examines the category from another point of view, congruent with Cruse’s notion. DiMarco et al. consider two dimensions of synonymy, connotation and denotation and, equivalently, semantic and stylistic differences. If two concepts differ semantically, that is in their denotation, such as *mist* and *fog*, they consequently cannot be considered synonymous. Moreover, as a substitution of one lexical item with the other would change the truth value of a statement, the

two categories can be considered to be plesionyms. Equally, words that only exhibit stylistic differences in register or vary only slightly in their denotation fall into the category of near-synonyms. While this distinction looks relatively clear cut and straight forward, the borders, as ever so often in linguistics, are fuzzy and the transitions fluent, often accompanied by overlapping areas. DiMarco et al. exemplify their claims concerning near-synonymy by using an example which illustrates that the two terms *woodchuck* and *groundhog*, even though denoting the same animal, cannot be used interchangeably as *groundhog* in the expression *Groundhog Day*, an American holiday, cannot be exchanged by *woodchuck* without loss or at least variance of meaning (1993: 120).

Recapitulating it can be said that there is consensual agreement among linguists on the fact that synonymy is a continuous category extending from no or zero synonymy to absolute synonymy. In between these two extremes of the scale plesionymy, or near-synonymy, and propositional synonymy can be located. While propositional synonyms do not change truth-values, share the same denotation and vary only slightly in connotation, plesionyms constitute very similar concepts, yet are not mutually interchangeable without altering truth conditions. While synonymy in general and propositional synonymy is a well if not the most investigated upon field of lexical relations in linguistics, plesionymy has not received equal attention from researches and has always been considered a subtype of synonymy until recently. Cruse (2002: 490) admits that it is the least explored category, presumably due to its complex and linguistically difficult character. For instance Storjohann (2009), in her paper on plesionymy, argues that it is not the shared semantic traits of plesionyms that are their defining properties. Instead she alleges that as a matter of fact the differing properties of near-synonyms are the more salient ones. Consequently Storjohann suggests that Cruse's categorization of plesionyms as a subcategory of synonymy seems to be satisfactory at first sight but a closer look suggests otherwise.

Murphy (2003) proposes a slightly different taxonomy of synonymy which nevertheless, in essence, is a mere refinement of the above. She tackles the issue of polysemy, the ability for a word or sign to have a number of meanings or signifieds. Within denotative sameness of meaning, which Cruse considered

absolute synonymy and propositional synonymy, Murphy (2003: 146) differentiates between two dimensions, namely how many senses a word has and how similar these senses are. According to her, full synonyms are identical in every sense and, not particularly surprising, words with a limited number of senses exhibit a higher tendency to be more synonymous than words conveying a plethora of senses. Murphy lists *woodchuck* and *groundhog* as full synonyms even though, as pointed out earlier, these two items cannot be considered absolute synonyms based on their context-dependent connotational dissimilarity. However, as at this point only denotation is factored into the analysis, *woodchuck* and *groundhog* can be considered full synonyms. These implications again reinforce the assumption that while absolute interchangeability of two lexical items is next to impossible, any subtype of synonymy potentially possesses the ability to replace another item in certain respects without loss or variance of meaning. Additionally Murphy subsumes words that distinguish themselves solely in stylistic nuances such as register under the category of full synonymy since, again, only denotational equivalence is considered. Interestingly enough she also mentions cross-language synonymy. As briefly discussed in the opening chapter, most linguists consider sameness of language a crucial property of synonymy and refuse the acceptance of language-blind synonymy. The second type of synonyms listed by Murphy is sense synonymy. Sense synonyms generally share some senses with their counterparts but differ in others. Both these types, full synonymy and sense synonymy are categorized as logical synonyms by Murphy. Therefore Murphy's taxonomy is a consequent development which enhances Cruse's notion of propositional synonyms, sets of words that do not change the truth-value of any given statement and share the same denotation but differ in certain connotational aspects. Propositional synonyms that share all denotational meanings are called full synonyms and those that share one or more but not all meanings are classified as sense synonyms. The example given by Murphy (2003: 147) are the words *couch* and *sofa*. Both words denote and share the meaning 'long upholstered seat' but the former also denotes 'a priming coat of paint', a sense not shared with its sense synonym.

Synonyms with one or more similar but not exactly the same senses that consequently change truth conditions when substituted, identified as near-

synonyms or plesionyms by Apresjan (1973) or Cruse (1986), are referenced as belonging to the hyperonymous category of context-dependent synonyms by Murphy (2003: 147). Words that share one or more senses which are not exactly identical yet very similar in meaning are defined as near-synonyms, consistent with other definitions of near-synonymy. Murphy then goes on to show that if there existed such a thing as two or more words sharing all senses which are similar but not exactly the same, a name for said category has not been determined yet. Murphy's taxonomy of the dimensions of synonymy is illustrated in Table 1.

Table 1: *Dimensions of synonymy*

	IDENTICAL SENSES (logical synonyms)	SIMILAR SENSES (context-dependent synonyms)
ALL SENSES	full synonyms	?
ONE(+)SENSES	sense synonyms	near-synonyms (plesionyms)

Source: Murphy (2003: 146)

As can be seen, synonymy is not just a matter of sameness of meaning but is also closely linked to and intertwined with antonymy and polysemy as well as hyponymy and hyperonymy.

Unfortunately, when it comes to the concept of synonymy, linguistic nomenclature is far from being unified and different researchers and linguists propose and use different technical terms and taxonomies. Moreover, even one and the same linguist may, owed to the findings of their own research or scientific progress in general, change their views and types of categorization. The boundaries between plesionyms and propositional synonyms are fuzzy and far from clear-cut and therefore allow the room for interpretation. DiMarco et al. assess that:

“Many of the semantic distinctions between plesionyms do not lend themselves to neat, taxonomic differentiation; rather, they are fuzzy, with plesionyms often having an area of overlap. For example, the boundary between *forest* and *wood* ‘tract of trees’ is vague, and there are some situations in which either word might equally be appropriate”. (DiMarco et al. 1993: 120)

An endless debate over a normative definition of synonymy and its various subcategories or variations, as it is often argued that especially plesionymy constitutes a category on its own, has been going on for a while now (cf. Sparck Jones, 1986; Cruse, 1986; DiMarco et al., 1993; Church et al., 1994; Murphy, 2003). Edmonds and Hirst’s paper on near-synonymy and lexical choice (2002) deserves special mention at this point as they note that in some of their early papers they followed Cruse’s nomenclature but opt for a more transparent terminology and therefore introduce a notion called granularity of representation (2002: 115-118). According to Edmonds and Hirst, Cruse’s (1986) definition of synonymy as words that are identical in their denotation and, if at all, only differ in their connotation or, in other, Cruse’s words, are identical in “central semantic traits” and distinguish themselves in “peripheral traits” (1986: 267), is too broad and needs revision. Edmonds and Hirst question the specifications of what counts as central traits and what is categorized as peripheral trait and additionally debate how much similarity central traits have to exhibit and how much dissimilarity for central traits is allowed within peripheral traits. Edmonds and Hirst’s proposed solution is to incorporate granularity into any and every analysis of synonymy as “any definition of near-synonymy that does not take granularity into account is insufficient” (Edmonds and Hirst, 2002: 116). Granularity, according to Edmonds and Hirst, is the level of detail used to describe a concept. A coarse-grained analysis is a simplified representation of synonymy and primarily concerned with denotation, the basic meaning of a lexical item and widely overlooks connotational properties. A fine-grained analysis on the other hand incorporates a high level of detail including connotational aspects such as register and can therefore provide a high level of information on the item under investigation and encode subtle differences. Simply put, a word can have a couple of words signifying similar concepts and on a coarse-grain level, used by

laypeople and dictionaries or thesauruses, these words can be considered as synonymous but a fine-grained analysis reveals that the concepts actually differ in certain aspects.

To put it in a nutshell, while there is an ongoing debate in the field of linguistics over certain aspects of synonymy, there have been established some cornerstones of the category within cognitive linguistic that most researchers acknowledge and adhere to. The empirical part of this paper provides a fine-grained corpus-based analysis, as suggested by Edmonds and Hirst, on the synonymy of the words *big*, *large* and *huge* and categorizes the three adjectives as well as locating them within the continuous spectrum of synonymy ranging from zero to absolute synonymy according to Cruse's taxonomy incorporating Murphy's amendments.

1.3 Motivation

Synonymy is considered to be the most studied lexical relation (Gries and Otani, 2010: 121), the lexical field of SIZE seems to be the default area of research when it comes to investigating paradigmatic relationships between lexical items and in doing so, virtually every inductive research method, from introspection to experimental elicitation through to corpus-based methodologies have been utilized. With that being the case, of what use for science in general and linguistics in particular could yet another corpus-based study on the synonymy of a set of words from the semantic field of SIZE be and why would one want to conduct yet another study of said kind. While at first glance the answer to these questions might seem difficult, a closer inspection of the matter at hand provides numerous reasons for further research on synonymy within the field of cognitive linguistics.

First of all, language is a "dynamic and constantly changing system" (Landsbergen et al., 2010: 363), which means that what might have been true some decades ago could be far from reality nowadays. Semantic change is inevitable and happens all the time as natural languages are in constant flux.

Additionally, corpus linguists argue that conceptual knowledge is dependent on our knowledge and beliefs and varies not only from person to person but also over time (Leech, 1981: 70). Said facts in connection with the notion of absolute synonyms to have no reason to exist in natural languages supported by many linguists (i.e. Clark, 1992; Cruse, 1986) alone would already be one important reason for the need of further research. *Big* and *large*, canonically considered as being a fine specimen of absolute synonymy might, in fact, have been more synonymous some fifty years ago than they are now.

Secondly, not only language is a dynamic and ever changing system, also the paradigm of cognitive linguistics is constantly changing and evolving. As mentioned in the previous chapter, the ongoing debate over paradigmatic relations between linguistic items, membership within relationship categories and general nomenclature is far from finished which, in turn, means that there still is the need for additional research in order to get to the gist of the matter and fully understand the character of lexical relations.

In addition to the above, while many of the more recent studies have already gone beyond the stage of solely applying an introspective methodology in order to investigate upon lexical relations, for instance quantitative elicitation, and some of the most recent ones have even applied corpus-based approaches in connection with additional software to process the obtained data, there is still plenty of room for improvement. Gries and Divjak declare in their 2009 paper on the analysis of linguistic relations that “in spite of the prominence of the term ‘usage-based’ currently enjoys in cognitive linguistic publications and in spite of the fact that some approaches explicitly touch their criteria in corpus-linguistic terms, there are a few truly corpus based approaches to polysemy and near synonymy” (2009: 60). “Technology has been the major factor in the growth of corpus linguistics” (McCarthy and O’Keeffe (2010: 5) and indeed technological progress and the advent of freely searchable, parsed corpora contributed significantly to the study of lexical relations. The ability to store and compute masses of data on relatively small home computers brought computational corpus linguistics to the homes of numerous cognitive linguists and laid the foundation for the advent of this new branch of linguistics. Therefore, while

corpus linguistics is a particularly fruitful and rapidly growing methodology not only within the field of semantic relations within the paradigm of corpus linguistics but throughout all linguistic branches, it is a relatively young methodology which means there are many areas of synonymy that have yet to be tackled utilizing computational linguistics and corpus-based approaches.

Fourth, just as Gries and Otani (2010: 126) remark, many corpus-based studies within the domain of lexical relations of SIZE focus on pairs of either canonical synonymy or antonymy such as *big/large*, *small/little* or *big/small* and *large/little*. The present study not only investigates the thought-to-be synonymous pair of adjectives *big* and *large* but also includes the word *huge* for the reason that it allows to put the results into perspective and, together with tests for statistical significance, assists in creating a taxonomy of the three lexical items within the continuum of synonymy.

Lastly, and this might be the most important point, even within the field of corpus linguistics, many studies only focus on one aspect of lexical relation instead of exploiting the richness of information corpora can provide and statistical software can process. Gries and Otani (2010: 127) argue that most corpus based studies on synonymy focus on either collocational aspects or syntagmatic relations exclusively but do not provide a comprehensive account of the category. Liu (2010: 58) favors corpus-based approaches to the study of synonymy because “corpus-based descriptions have been shown to be much more accurate and informative than traditional non-corpus-based descriptions” (2010: 58). Furthermore, he claims that the behavioral profile approach has been proven to be an extraordinarily adequate methodology for the study of synonymy as behavioral profiles, or BPs, factor in a wealth of distributional characteristic instead of just analyzing syntagmatic or semantic relations. To some extent the behavioral profile approach is based on the work of Firth who was the first linguist to establish collocational analyses as central concept of lexical semantics (Liu, 2010: 58). Probably every linguist has, at some point, come across Firth’s famous and groundbreaking assertion that “the complete meaning of a word is always contextual” (1957: 7), which paved the way for an entirely different conception of synonymy within linguistics. To “know a word by the company it

keeps” (Firth, 1957: 11) in connection with “knowledge of how that word is used” (Miller and Charles, 1991: 4) is, in principle, what constitutes a behavioral approach to synonymy. Syntactic considerations paired with semantic and pragmatic properties provide a full-fledged account of the lexical relationship of two or more concepts. As established previously, corpus linguistics is a relatively recent field of linguistics. Nevertheless, behavioral profile analysis is an even younger field within a corpus-linguistic framework without which BP analysis would not even be possible. This fact obviously means that there are various unexplored areas which call for additional, further research.

To sum up, while it is true that linguistic categorization in general and synonymy, in particular, especially within the semantic field of SIZE are, in fact, a relatively well established and investigated upon area in linguistics, there are a number of reasons that call for further research. Firstly, the advent of electronic corpora yielded new, more all-encompassing methodologies better suited to provide comprehensive accounts of synonymy. Secondly, language change and paradigm shifts also affect linguistic research and, last but not least, the infancy of corpus-linguistics and behavioral profile analysis provide sufficient grounds for the study at hand.

1.4 Theoretical framework and previous studies

Due to the fact that the paper at hand constitutes a corpus-based behavioral profile approach to the investigation, this chapter briefly discusses general, non-corpus-based approaches to the study of synonymy which can be considered the foundation for the more recent corpus-studies on synonymy which subsequently yielded corpus-based behavioral profile analyses.

Generally, “the issue of synonymy and antonymy has mostly contrasted two different perspectives: the co-occurrence approach (cf. e.g. Rubenstein and Goodenough 1965) and the substitutability approach (cf. e.g. Deese 1962, 1964)”, according to Gries and Otani (2010: 122).

The co-occurrence approach is based on collocational properties of linguistic items and can be traced back to influential linguists such as Firth (1957) and Bolinger (1968) who were in favor of syntagmatic analyses as their approach to the meaning of the word highly correlated with the affinity of any given lexical item with its linguistic environment. Bolinger is quoted by Gries and Otani (2010: 122) declaring that “a difference in syntactic form always spells a difference in meaning” (1968: 127). Many other linguists besides Bolinger and Firth hold the opinion that the co-occurrence of lexical items is closely linked to their usage and therefore their meaning. These include but are not limited to Cruse (1986), Hanks (1996) or Harris (1970). Their consensual theoretic assumptions then paved the way for a variety of empirical studies on synonymy. In the very beginning, non-corpus-based methodologies such as introspection or elicitation through tests and questionnaires were utilized, later on supported by electronic data. Notable corpus-based studies adhering to the co-occurrence approach and thus providing collocational and syntactic information include Church et al. (1991) and their investigation on which words are modified by *strong* and *powerful*, Atkins and Levin (1995) on the grammatical associations of *little* and *small* and, most relevant for the study at hand, Biber et al. (1998) and their findings concerning the distributional patterns of *big*, *large* and *great*. Biber et al. found out that out of a 5.7-million word sample from the Longman-Lancaster corpus, *large* was the most common adjective with a normed frequency of 408 occurrences per million words, followed by *great*, 383 instances and then *big* with a frequency of 203 tokens (Biber et al., 1998: 43). Frequency distributions by genres for academic prose and fiction reveal that *large* is particularly common in academic prose with 605 normed instances whereas *big* is highly uncommon with only 31 occurrences per million words, a fact that already suggests that even *big* and *large*, the two lexical items canonically considered to exhibit both denotational and connotational synonymy are, in fact, not absolute synonyms. The findings of the study are depicted in table 2 which already shows from simple, raw frequency counts that the adjectives *big*, *large* and *great* are not evenly distributed among genres and consequently such a kind of variation in the usage of the lexemes is indicative of a difference in meaning.

Table 2: Raw frequencies of *big*, *large* and *great* in a 5.7-million-word sample from the Longman-Lancaster corpus

	Raw counts	Normed per million words
Total Sample (5.7 million words)		
<i>big</i>	1,319	230
<i>large</i>	2,342	408
<i>great</i>	2,254	393
Academic Prose (2.7 million words)		
<i>big</i>	84	31
<i>large</i>	1,641	605
<i>great</i>	772	284
Fiction (3 million words)		
<i>big</i>	1,235	408
<i>large</i>	701	232
<i>great</i>	1,482	490

Graphic representation of normed frequencies (each * represents 100 occurrences per million words)

	combined	academic	fiction
<i>big</i>	**	-	****
<i>large</i>	****	*****	**
<i>great</i>	****	***	*****

Source: Biber et al. (1998: 44)

Additionally, Biber et al. learned that *big*, throughout both genres, almost exclusively collocates with words that refer to physical size, making it a monosemous concept, whereas *large* and *great* exhibit a number of other, polysemous, metaphorical usages. This, again, hints towards a rejection of absolute synonymy between the terms *big* and *large*.

Substitutability approaches, according to Gries and Otani (2010: 124), are conducted as follows. A number of sentences containing a linguistic item are collected from a corpus or, especially in the pre-electronic era, elicited from subjects in a study. Subsequently, a number of sentences containing the alleged synonymous expression are gathered. The items then are deleted from the sentences and the sentences, at random obviously, are presented to subjects who are supposed to decide which one of the items is a better fit. The more sentences can, according to the test subjects, take both expressions, the more similar and therefore synonymous the items are judged to be. Notable proponents and studies belonging to the substitutability approach are Charles and Miller (1989) or Rubenstein and Goodenough (1965).

When it comes to previous research following a substitutability approach in connection with the lexical field of SIZE, one of the first studies containing an empirical part was Deese's (1964) paper on the associative structure of some common English adjectives. Deese analyzed the 278, according to Thorndike and Loge (1944), most frequent adjectives in the English language with regard to their distributional patterns. Analyzing the results of a stimulus-response word association test he concludes that *big* and *large* are very similar in meaning while *great* scores a slightly higher rating. Although a substantial part of the study is dedicated to antonymy, his findings concerning synonymy are considered to be an important landmark for the study of lexical associations.

These theoretical frameworks molded by philosophers and linguists from the 1950s onwards were then, with the advent of corpus linguistics, enhanced and empirically tested. While, as mentioned above, many of the early corpus-based studies on synonymy heavily relied on one or only a few aspects to be taken into consideration when analyzing lexical associations, the implementation of

behavioral profile analyzes yielded a new, holistic approach to the study of lexical relations. In their 2010 study, Gries and Otani advocate the sheer infinite possibilities and the wealth of data a behavioral profile analysis has to offer (2010: 142). They claim that the main differences between the BP approach and their former yet, as they admit, similar precursors are of a twofold character. First of all, the BP approach takes into consideration not only one dimension of synonymy, syntactic or semantic relations, but analyses both properties in a fine-grained way. Moreover, the range of contextual or distributional characteristics is higher because instead of just looking at a single element right of the item under investigation for attributive adjectives, for instance, the BP approach includes a plethora of linguistic features from different levels. Syntactic constraints, syntagmatic relations, collocational properties or raw and normed frequencies are only a few of the types of information a BP approach can deliver and the consideration and incorporation of all of these bits are what makes the BP approach unique and superior to earlier, simpler approaches. As for the results of Gries and Otani's study (2010), it can be said that besides the confirmation of the canonical antonym pairs *big/little* and *large/small* they infer, from manually annotated and quantitatively analyses ID tags and the corresponding dendrograms and snakeplots that *big* is more similar to *large* than great. Furthermore, when comparing *big* and *large*, the relative small deviations from 0 show that the two items are generally highly synonymous yet, by means of fine-grained analysis, it can be seen that there are some distinct differences. *Large* is more prone to modify count nouns that refer to quantities, organizations and animate entities excluding humans. *Big*, on the other hand, exhibits a slightly different behavioral profile. It tends to rather modify non-count nouns and co-occurs to a higher degree with abstract nouns, humans and actions. The lexemes also exhibit slight differences in their usage concerning tense and mood of the environment they are used in (2010: 141). Although Gries and Otani have carried out a very well conducted study and their findings mostly confirm previous assumptions and outcomes, they report that particularly their take on attributive and predicative predispositions of certain adjectives contradict findings reported by Biber et al. (1998), especially considering that both studies employed the same corpus and data. However, they conclude their most relevant findings

conform to the outcomes of Biber et al. The snakeplot for *big* and *large* in Gries and Otani's study is depicted in Figure 1.

Figure 1: Snakeplot of differences between BP vector values: *big* – *large*

Source: Gries and Otani (2010: 140)

1.5 Aim of the paper and hypothesis

Based on the supposition that, from a cognitive point of view, actual synonymy is beyond rare, it is to be assumed that the adjectives *big*, *large* and *huge* share some linguistic features but differ in others. As a matter of fact, previous studies have already established an extensive body of work on the synonymy of *big* and *large*, notably Gries and Otani (2010) or Biber et al. (1998). The study at hand however utilizes the Corpus of Contemporary American English, an even more all-encompassing corpus of the English language than the corpus employed in Gries and Otani's study, for instance. Additionally, while it is not to be expected that the data gathered from the Corpus of Contemporary American English contradicts the findings of other studies, its richness of information in connection with a very fine grained analysis will provide even more information on the fine shades of semantic and syntactic differences of the two items. Moreover, the word *huge* has been widely neglected in most studies on synonymy which, to this date, mainly focused on the canonical pairs *big* and *large*. The inclusion of *huge* will allow for a placement of the three adjectives on the continuous scale of sameness of meaning and put the close proximity of *big* and *large* into perspective. It is hypothesized that *big* and *large* exhibit ample areas of linguistic overlap and may, in many respects, be considered to be absolute synonyms. *Huge*, while still featuring and sharing many vital properties with the others, offers some distinct characteristics making it less synonymous but in certain aspects and environments, identified in the study - however distinct in general - it can nevertheless be identified as cognitive synonym of *big* and *large*. Based on Biber et al. (1998), who claim that *big* almost exclusively modifies lexical items belonging to the semantic field of SIZE whereas *large* shows a number of other, metaphorical uses, it is assumed that the overall number of occurrences for *big* is higher than for *large* and *huge*, which is also considered synonymous to *big* and *large* by most thesauruses, shows the least number of tokens due to its alleged restrictedness. Furthermore, the paper at hand divides its analysis up into the five genres found in the Corpus of Contemporary American English, namely academic, newspaper, fiction, magazine and spoken, and analyzes whether one of the three adjectives investigated upon is more prone to be used in a certain

genre and proposes why this could be the case. This is another, never examined before and therefore relevant, aspect of the paper as Biber et al. (1998) have only considered two genres, fiction and academic prose.

To sum up, it is assumed that the data gathered from the Corpus of Contemporary American English will, for the most part, confirm and reinforce previous studies. Additionally, further aims of the paper are to provide an even more extensive account of the synonymy of the words *big* and *large* through very fine grained analyses and representative samples of state-of-the art data. The adjective *huge*, mostly neglected in previous studies even though listed as synonym of both *big* and *large* in various dictionaries and thesauruses, is incorporated into the study and checked for relative synonymy regarding some meanings of its polysemous nature. Finally, the findings are checked against and categorized according to Cruse's (1986) framework of truth-values and Murphy's (2003) amendments concerning polysemy and partial synonymy. This study aims to go beyond, at least in some respect, what some previous works have done in so far as it investigates the distributional behavior of the set of adjectives belonging to the field of SIZE from a number of different points of view including semantic relations and syntactic features instead of simply focusing on just collocations or raw frequencies. Lastly, multiple correspondence analyses provide graphic representations of the findings, a powerful tool not utilized in previous studies.

2 The study

In the empirical part of the thesis, the source of data, methods utilized, annotation and data filtering routines as well as additional software are presented and explained. The results of the study are depicted and discussed in terms of validity, limits and theoretical ramifications.

2.1 Source of data and choice of method

The study at hand is a corpus-based study on synonymy within the theoretical frameworks and paradigms of cognitive linguistics. The main assumptions of cognitive linguistics revolve around three main aspects. First of all, general cognitive processes are responsible for language production and comprehension instead of an innate, autonomous entity in the mind. Secondly, knowledge of language is knowledge of how language is used and the usage of language is acquired and contextual, as opposed to invariable and independent of speaker as in autonomous linguistics. Thirdly, a difference in surface structure, or grammar, correlates with differences in deep structure or conceptualization (Croft and Cruse, 2004). These three main concepts partly determine or at least favor certain methodological considerations. The notion that grammar is the manifestation of conceptualizations of speakers of languages determines that syntactic aspects are essential to the analysis of linguistic concepts and thus behavioral profile analysis cannot omit syntagmatic relations. Moreover, the assumption that knowledge of language is knowledge of how the language is used asks for a quantitative analysis of large samples of actual language production in order to draw conclusions for the language in general as opposed to qualitative analyses of individual speakers.

Electronic corpora combined with additional software for statistical computing and optional graphic representation allow for exactly these tasks to be fulfilled. Annotated corpora provide huge chunks of data readily available and examinable in terms of syntactic properties, semantic relations and collocational patterns. The

few not readily retrievable characteristics can be annotated manually and processed by statistical software. Consequently, a corpus-based approach is applied to the study at hand.

Various different types of corpora are in existence, each distinct type serving different purposes. The most distinguishing feature however is if the data is electronically stored or not. Today's linguists, when they refer to a corpus almost exclusively talk about electronic corpora. According to Kennedy (1998: 3), for instance, a corpus is "a collection of texts in an electronic database that serves as the source of corpus linguistics". There has been quite some debate, especially in the early years of corpus studies, if corpus linguistics constitutes a methodology within linguistics suitable for all branches or if corpus linguistics should be considered a separate branch of linguistics on its own. Kennedy perceives corpus linguistics with its evidence derived directly from actual speech production to be an enhancing methodology instead of considering it a fully-fledged branch of linguistics (1998: 3-5). Kennedy also confirms what has already been addressed in earlier chapters of this thesis and what is immediately evident to most linguists. Researchers have always needed sources and linguistic data to formulate and substantiate theories and electronic corpora that possess the ability to store and make available sheer endless amounts of linguistic data make it much easier to find, identify, count or sort linguistic items. McEnery and Wilson (2001: 7) raise yet another point in favor of corpus linguistics. According to them, "linguists are often more interested in a whole variety of language rather than an individual text" and indeed, when it comes to trying to analyze language use and language as a whole, large non-specialized corpora are as representative of actual language production as possible. The mere inexhaustibility of modern corpora in addition with their constant expansion and equip corpus linguists with a huge advantage over conventional methodologies in this very respect. Fields of application, according to McCarthy and O'Keeffe (2010: 5) include but are not limited to language teaching and learning, discourse analysis, literary and translation studies, forensic linguistics, pragmatics or sociolinguistics.

All these aspects suggest that an extensive electronic corpus serves as the best source in order to tackle the task of providing an all-encompassing behavioral

profile analysis of the synonymy of *big*, *large* and *huge*. However, with the abundance of different types of corpora available extreme caution has to be exercised when it comes to picking the most appropriate one. First of all, a number of corpora have already been utilized for studies on synonymy and therefore, even when considering different syntactic and semantic aspects, there is no use in simply recreating an already existing study. Furthermore size and coverage of the corpus have to be considered. Due to its size, its powerful in-built concordancer and its internal structure, the Corpus of Contemporary American English is the database of choice for the study at hand.

2.1.1 The Corpus of Contemporary American English

The Corpus of Contemporary American English is one of the largest and most up to date corpora of English. It is a 450-million-word corpus and thus more than 4 times the size of the British National Corpus, a comparable corpus of British English containing approximately 100 million words. The COCA, as the Corpus of Contemporary American English is abbreviated, is also incomparably larger than the 5.7-million-word sample of the Longman-Lancaster corpus utilized by Biber et al. for their 1998 study. It has been created by Mark Davies, professor of linguistics at Brigham Young University. Its size renders the corpus the most representative sample of general language available. With its most recent update in the summer of 2012, it is also the most up to date corpus of English which serves the purpose of assessing whether there might have occurred a shift of meaning in either denotation or connotation of one of the three adjectives under investigation in between earlier studies and the one at hand. Moreover, the COCA is divided up into five different genres allowing for investigations upon synonymy and possible varying linguistic relationship within these categories. The powerful and sophisticated in-built concordancer paired with the parsed and tagged for grammatical information and parts of speech language samples allow for precise searches and provide exact results. The interface not only allows users to search for exact words or phrases but also accepts wildcards and can be employed to search for parts of speech such as adjectives just like in this very study, or lemmas and any combinations of the two above. As an added benefit,

the COCA is searchable for collocates within a ten-word window which is, as already mentioned, a central and crucial aspect of behavioral profile analyses (Davies, 2008-). All these characteristics combined make the COCA the best choice for a representative, all-encompassing and scientifically valid study on synonymy.

2.1.2 Behavioral profile analysis

The data collected from the corpus is then processed in terms of a behavioral profile analysis. As already established, when working on synonymy linguists need ways to somehow measure the degree of similarity in meaning of words or concepts under investigation and one possibility is a relatively recent innovation called behavioral profile analysis. Divjak and Gries are two pioneering linguists in the field of BP analyses (Divjak 2006, Divjak and Gries 2006, Divjak and Gries 2008, Gries and Divjak 2009). In their 2009 paper on BP approaches, Gries and Divjak introduce and comment on a number of other approaches typically employed when investigating synonymy or the very closely related concept of polysemy. Such approaches include the so called full-specification approach by cognitive linguistics' grey eminence George Lakoff (1987) or the related theory of a partial-specification approach by Kreitzer (1997). Both these experimental frameworks however, according to Gries and Divjak (2009: 57), suffer from "methodological inadequacies and representational problems". Gries and Divjak cavil that both the full-specification approach as well as the partial-specification approach fail to properly take the context of the word under study into account. This is highly interesting due to the fact that Lakoff is one of the founding fathers of cognitive semantics and context-dependency is one of the main assumptions of said branch of linguistics. Tyler and Evans (2001) argue in a similar way to Gries and Divjak. Another approach taken up by Gries and Divjak (2009: 58) is Sandra and Rice's (1995) method which, much like behavioral profile approaches, utilizes a variety of experimental settings in order to obtain a more comprehensive and objective account of linguistic relations. Test subjects were asked to cast sentence acceptability judgments, sentence similarity judgments and generate sentences on their own. However, Gries and Divjak remain

skeptical towards the approach and list a number of misconceptions such as the question of whether conscious classification actually reflects underlying mental representation (2009: 58). They conclude that despite the popularity of the term ‘usage-based’ within the cognitive framework, only but a few studies actually utilize a usage based methodology. The one notable exception Gries and Divjak mention is the corpus-based study of Kishner and Gibbs (1996) on *just*. Nevertheless even Kishner and Gibbs’ study, which investigates colligations and R1 collocations, does not fully exploit the possibilities offered by corpora, among others due to the small collocation window of only one position. In comparison, the Corpus of Contemporary American English is capable of examining collocates within a ten-word-window. As a consequence of the shortcomings and failures of most non-corpus-based studies and the lack of fully exploiting the potential of corpus linguistics, inherent to many of the early usage based studies, Gries and Divjak advocate an approach known as behavioral profile analysis. The behavioral profile approach is a multifactorial analysis of linguistic traits and as a corpus-based method it customarily starts out by doing what electronic corpora do best, counting frequencies. The assumption behind frequency counts is that “distributional similarity reflects, or is indicative of, functional similarity” (Gries and Divjak, 2009: 61). If the distributional pattern shows an uneven distribution which is, verified through statistical tests, stronger than what could be attributed to mere chance, this is already a hint towards non-absolute synonymy a further investigation on the determining reasons is needed. Gries and Otani (2010: 128) propose four major steps for the application of the BP method which, in essence, coincide with the points put forward by Gries and Divjak (2009: 64). The four steps identified by Gries and Otani read as follows:

- (i) the retrieval of (a representative random sample of) all instances of the lemmas of the synonyms/antonyms to be studied from a corpus in the form of a concordance;
- (ii) a (so far largely) manual analysis and annotation of many properties of each match in the concordance of the lemmas; these properties are, following Atkins (1987), referred to as ID tags and include morphological, syntactic, semantic, and collocational characteristics;

- (iii) the conversion of these data into a co-occurrence table that provides the relative frequency of co-occurrence of each lemma with each ID tag; the vector of these co-occurrence percentages for a lemma is called that lemma's behavioral profile;
- (iv) the evaluation of the table by means of exploratory and other statistical techniques, especially hierarchical agglomerative cluster analysis.

These steps serve as general outline of a corpus-based behavioral profile analysis and their application and implementation is presented in section 2.3, data selection, filtering and annotation.

2.1.3 Dictionaries and thesauruses

In order to provide an account of the concept of synonymy in lexicography in the introductory chapter, four selected dictionaries have been surveyed, namely the Oxford Online Dictionary, the Collins English Dictionary, the Oxford English Dictionary and the Concise Oxford Dictionary of Linguistics. Additionally these dictionaries have been evaluated in terms of meanings of *big*, *large* and *huge*. It has been established whether the dictionaries list one or both adjectives as synonyms to their respective counterpart. Moreover, it has been determined if the adjectives are listed as polysemous items and if so what their proposed meanings are so as to compare the entries with the findings provided by the Corpus of Contemporary American English.

The Oxford English Dictionary, abbreviated as OED, is a multi-volume historical dictionary created and published by the Oxford University press. Work on it began as early as 1857 and the fact that it contains approximately 750,000 words makes it the world's most comprehensive dictionary. It is considered a, if not the standard reference for linguists of all sorts (Winchester, 2003).

The Concise Oxford Dictionary of linguistics is, according to Oxford University press, the most authoritative dictionary of linguistics. Its makers claim that it covers every aspect of the multidisciplinary field that is linguistics from phonetics

via semantics through to language theory. The author of the dictionary, Peter Matthews, is professor of linguistics at the University of Cambridge.

The Collins English Dictionary is an online and print dictionary which uses the *Bank of English*, a contemporary corpus of mainly British English. However, American and Australian texts are also included. The corpus has been developed by COBUILD, a research facility at the University of Birmingham. The acronym stands for *Collins Birmingham International Language Database*. The fact that the CED is based and mainly provides examples from the corpus suggests that its definitions and examples are, as they are actual language products and used by native speakers, very close if not identical with the results of the study.

2.2 Supporting Software

In order to be able to process, graphically depict and interpret the data extracted from the corpus, additional software is needed. Step two of the behavioral profile analysis asks for a manual annotation and creation of ID-tags. For that purpose a spreadsheet was used. Subsequently, in the next step, the data is converted and fed into the language R environment for statistical computing.

2.2.1 Language R, statistical computing

The R project, a programming language for statistical computing, is capable of processing different types of quantitative data, able to carry out various statistical tests such as chi-squared tests and can be used to plot and graphically depict the results. Language R is a free software environment for statistical computing and graphics very similar to IBM's SPSS software with the advantage that R is an open source project and therefore free of charge. R is a GNU-project which guarantees the end user the freedom to use, study, modify and share the software by definition. It compiles and runs on a wide variety of platforms including Windows and MacOS.

2.3 Data selection, filtering and annotation

In the first step, retrieval of all instances of a word (Gries and Otani 2010: 128), all occurrences of the adjectives *big*, *large* and *huge* were retrieved utilizing COCA's in-built concordancer. As the paper at hand is only concerned with the adjectival uses of the words, other possible parts of speech, such as adverbial uses as in (7), are excluded from the study.

- (7) Will just talks big. (COCA: Lund, Gerald: *Pillar of Light*. Fiction. Salt Lake City: Bookcraft, 1990)

The search string “word.[pos]” provides the exact word, as this study is not concerned with all the lemmas for *big*, *large* and *huge* but only their base forms, as part of speech. Therefore, “big.[j]” yields all instances of the adjective *big* to be found in the Corpus of Contemporary American English. Thereafter, for project feasibility, a statistically relevant subset has to be selected. A statistical power analysis determined the necessary number of tokens per type for a certain probability to find the desired effect, if it exists, within a particular confidence level. The calculation was conducted for an 80 percent probability of finding the effect within a significance level α of 0.05 percent or a confidence interval of 95 percent. The next step involved a manual annotation of the lexical items in the subsample. These annotations are called ID-Tags by Gries and Otani and their range is almost endless as any linguistic category at any linguistic level can be included, one of the factors that distinguish the multifactorial behavioral profile analysis from conventional, mainly monofactorial approaches. The number of factors included and the linguistic levels included determine the level of granularity of a study. Due to the fact that this study is conceptualized as a very fine-grained analysis and an amendment to previous studies on the sameness of meaning of the canonical synonym pair *big* and *large* and their potential sense synonym *huge*, a variety of linguistic categories is implemented and analysed within a 10-word collocation window. Table 3 lists the linguistic properties, or ID-Tags, considered in this study.

Table 3: ID-Tags for the behavioral profiles of big, large and huge

Type of ID-Tag	ID-Tag or variable	Level of ID-Tag or variant
Morphological	<i>tense</i>	<i>present/past/future</i>
	<i>voice</i>	<i>active/passive</i>
Syntactic	<i>sentence type</i>	<i>declarative/interrogative</i>
	<i>clause type</i>	<i>main/dependent</i>
	<i>type of adjective</i>	<i>attr./predic./postpos.</i>
Semantic	<i>type of object</i>	<i>concrete/abstract</i>
	<i>object</i>	<i>animate/inanimate</i>
	<i>countability</i>	<i>countable/non-countable</i>
	<i>negation</i>	<i>present/absent</i>
	<i>SIZE modification</i>	<i>literally/metaphorically</i>
	<i>connotation</i>	<i>positive/negative/neutral</i>
	<i>semantic field</i>	<i>technology/leisure/...</i>
Lexical	<i>collocates</i>	<i>collocate₁, collocate₂, ...</i>

Source: Modified version of Table 1: Selective overview of ID-Tags and their levels
(Gries and Divjak, 2009: 65)

Table 3, a modified version of Gries and Divjak's (2009: 65) table summarizing ID-Tags which have been used in behavioral profile analysis, outlines the general annotation scheme for this study. Starting from the premise that *big*, *large* and *huge* are actual synonyms, variability between the number of variants for each item should not be higher than what tolerance levels of normal distributions allow. The numbers are statistically tested using a chi-squared test and its outcomes determine whether distributional patterns vary because of some linguistic effect

associated with the adjectives or due to mere chance. While the annotation of syntactic and morphological aspects is generally relatively straightforward as there are prescriptive grammatical rules to be followed and these categories are fairly clear cut, ID-tagging for semantic properties is not always as simple and individual tags may be subject to discussion.

The first variable, tense, is divided up into present, past and future and regards the finite verb of the sentence or clause in which the adjective is used. The same is true for the ID-Tag voice and its variables active and passive. As for the syntactic variables, sentence type determines whether the adjective occurs in a declarative or interrogative sentence, clause or utterance. Clause type is split up into main or dependent. The category type of adjective classifies the type of adjective to be found and consists of the categories attributive, predicative and postpositive. Semantically, type of object refers to the noun the adjective modifies and is divided up into concrete and abstract and the category object determines whether the referent is animate or inanimate. If an adjective is categorized as countable or uncountable depends on the noun it enhances and its respective properties. Moreover, presence or absence of negation within the utterance the adjective is used in is assessed. One of the most important aspects for the analysis is the question of whether the adjective under investigation modifies the semantic field of SIZE literally or metaphorically. Lastly, connotation, judged by whether the adjective is used in a positive, negative or neutral lexical and semantic environment or context, is determined. In addition to all of the above, all tokens of *big*, *large* and *huge*, are assigned to one of the five genres within the COCA, namely spoken, fiction, magazine, news and academic. To exemplify the taxonomy incorporated into the study, a sample sentence is presented in (8).

- (8) Then there's the offense, which fizzled after building the **big** lead, and the defense, which gave it away, and finally, then-coach Jim Fassel, who could have been a little more assertive about the missed call and a lot less conservative in his play-calling. (COCA: Litke, Jim: *It brings the taste of vomit back*. News. Associated Press, 2012)

Classified according to the annotation criteria this example would thus be listed

as past tense, active voice, declarative statement, dependent clause, attributive, abstract, inanimate, uncountable, no negation, metaphorical modification and negative context. The annotated data is then, in a further step, entered in a spreadsheet software in order to be compatible with language R's quantitative analysis routines. An excerpt of such a so-called co-occurrence table, based on Gries and Divjak (2009: 66) is Table 4, depicted below.

Table 4: Excerpt co-occurrence table *big, large and huge*

Citation	Tense	Voice	Sentence Type
<i>Then there's the offense, which fizzled after building the big lead, and the defense, which gave it away...</i>	<i>past</i>	<i>active</i>	<i>dependent</i>
<i>So what you have is a great big intelligent piece of iron without a crew, " he said.</i>	<i>present</i>	<i>active</i>	<i>main</i>
<i>Coronary heart disease (CHD) is big business for the NHS.</i>	<i>present</i>	<i>active</i>	<i>main</i>

Source: Author's illustration (2014)

In a further step, this table is then converted in frequencies by language R and can then be analyzed quantitatively. The according statistical tests and the conversion into graphical depictions and their interpretation are explained in the following chapter.

2.4 Choice of statistical tests and graphical representation

Behavioral profile analyses are inherently quantitative in nature. Even though they raise the claim to holistically explore and explain a phenomenon, in order to do so, they necessarily need to be enhanced by qualitative considerations, just like the one at hand. Due to its quantitative aspiration, this study, like virtually any corpus-based study, begins with “counting examples of a certain kind and comparing them to another” (Glynn, 2010: 11). By means of computational linguistics and with the aid of corpora, linguists are able to retrieve huge chunks of data which, in further consequence, have to be processed statistically. Gries (2010: 269) argues that corpus linguistics is a distributional discipline by its nature. “Corpora as such contain nothing but distributional frequency data”, he argues. As a matter of fact, even the most elaborated and up to date corpora heavily rely on frequency information but, thus far, lack additional linguistic information which is one of the reasons why behavioral profile analyses require manual annotation of the subsample under investigation. Glynn strongly advocates statistical analyses of corpus data as he claims that modern linguists are confronted with an increasingly multi-faceted field of study and even more so with increasingly multifactorial methods and are therefore dependent on powerful analytical tools (2010: 12).

Beside the conventional frequency analysis, raw and standardized, and collocational searches carried out as starting point of this and virtually any corpus-based study, a number of statistical tests have been employed. First of all, prior to collecting any data, a so-called statistical power analysis was conducted (cf. Cohen: 1988; Butler: 1985) to plan the project. A power analysis is utilized to estimate the necessary sample sizes in order for it to be representative of the entire population in connection with a certain possibility of finding the desired effect. Subsequently, after retrieving an appropriate number of data from the corpus, chi-squared analyses for goodness-of-fit in order to check whether the distribution of the adjectives is not equally but randomly, distributed and thus no effect can be located, have been conducted. A chi-squared test for independence and an analysis of variance was incorporated so as to find out

whether the variable genre has an impact on the distribution of *big*, *large* and *huge*.

Additionally to the aforementioned standard statistical tests, another type of analysis has been implemented in the study at hand, the multiple correspondence analysis. Other types, such as the dendrogram or the snakeplot, have been spared as they have already been utilized in a number of similar studies and one of the aims of this paper is to either, within certain categories, reinforce the notions about the synonymy of *big*, *large* and *huge* or, within other areas, to falsify particular assumptions introduced by previous studies. The most relevant studies which utilized dendrograms or snakeplots are covered in section 1.4 on previous studies. Abdi and Valentin 2007: 652) define a multiple correspondence analysis as a means which allows to “analyze the pattern of relationships of several categorical dependent variables”. It is mainly used to analyze sets of nominal variables which makes it the test of choice for the purpose of the paper at hand. Correspondence analyses convert the frequencies of multiway tables into physical distances plotted in a two-dimensional graph (Glynn, 2010: 12). This graph depicts the multi-faceted correlations and differences between the categories visually which, in turn, allows for interpretation and explanation. Nevertheless, Glynn (2010: 12) argues, sometimes the interpretation of these two-dimensional graphs formed from n-dimensional correlations and differences can be tricky and subject to misinterpretations. Caution should be exercised as “the position of the data points relative to other points can be misleading” and “careful consultation and experience interpreting the plots is the only way to avoid misinterpretation”, Glynn (2010: 12) advises. As mentioned, a language R correspondence analysis converts frequencies into distances and consequently relative proximity of a data point to another means relative correlation. In other words, the closer two data points to one another, the more often they share a certain trait in the samples taken and therefore, assuming that properly chosen and sized samples are representative of the general population, actual language. This is a very important insight as a multiple correspondence analysis of that type can deliver certain aspects, for example syntactic considerations, in which two of the adjectives are always interchangeable, thus absolutely synonymous in this

respect, and some areas where they exhibit no congruency whatsoever. Obviously, with *big*, *large* and *huge* being proven as relatively synonymous, the areas of linguistic overlap will outweigh those of non-congruence by far yet this is an important aspect for the implementation of the findings of the study in teaching for instance. The most flamboyant properties of the three adjectives and their differences and similarities in usage concerning syntactic structures and semantic fields are worked out in the study in can, in further consequence, be used by teachers in order to teach learners of English as an L2 the usage of *big*, *large* and *huge* as done by native speakers. Lastly, to statistically support the interpretations of the multiple correspondence analyses, an agglomerative cluster analysis has been carried out so as to obtain approximately unbiased p-values and bootstrap probability values to mathematically support the interpretations.

To put it in a nutshell, if done correctly, statistical tests and the visualizations of correspondences provided by R are very powerful tools. The power analysis facilitates project planning, chi-squared tests ensure scientific validity of the results and minimize logical fallacies as well as inferential errors, the cluster analysis supports the data and the correspondence analysis allows for interpretation and explanation of the results.

3 Results

Excerpts of the definitions for the three adjectives under investigation, *big*, *large* and *huge* provided by the same four dictionaries consulted for the lexicographical definition of synonymy in chapter one of this thesis are presented in what follows.

3.1 Lexicographical analysis

The Oxford Online Dictionary, (OOD), defines *big* as “[o]f considerable size or extent” or “[o]f considerable importance or seriousness” and lists a number of sample sentences for both categories. Furthermore it lists a number of synonyms for *big*, among them, which comes as no surprise, *large* and *huge*. In essence, the two definitions provided by the OOD could be subsumed as the literal and metaphorical variants of the variable SIZE modification.

The Collins English Dictionary, (CED), offers a number of definitions for the adjective *big*, the most important ones are:

- “of great or considerable size, height, weight, number, power, or capacity”
- “having great significance; important”
- “important through having power, influence, wealth, authority, etc”
- “(intensifier usually qualifying something undesirable)”
- synonyms: large, great, huge, giant, massive, vast, ...

Similar to the OOD, the CED lists *large* and *huge* as synonyms of *big*. While at first glance the definitions one and two listed by the CED are relatively similar to those offered by the OOD, the CED does not make a distinction between the literal meaning of *big* and its modification of physical size but, different to the OOD, includes metaphorical modification in all its definitions.

The Concise Oxford Dictionary of Linguistics, as a linguistic dictionary, does not offer an entry for the adjective *big* and links to its sister dictionary the Oxford Online Dictionary, which is presented above. The Oxford English Dictionary, the

largest and most comprehensive general dictionary of the English language, offers the most information on the adjective. The fact that the OED contains dates for the first attested usages of words as well as etymological information in connection with its chronological presentation of changes in usage of lexical items, make it highly relevant for the task at hand. According to the OED, the first attested sense of *big* is, for animate objects, animals or humans, to have great strength. This sense however, as claimed by the OED, is obsolete. The corpus analysis will reveal if there are still instances that adhere to one of the adjectives' early senses. The second sense of the word recorded as early as 1400, is for humans to be rich or wealthy. This sense is classified as rare which, again, will make it interesting to see if the adjective *big* is still used in this sense which will be determined by the variable semantic field. The third sense addressed by the OED is marked as the principal sense nowadays. However, they add that in early use the meaning often overlaps and is difficult to distinguish from the two aforementioned senses and therefore setting an exact date for the emergence of this sense is virtually impossible especially with hindsight to the interpretational character of such a category. The adjective is defined as being "of considerable size, bulk or extend" (OED). Additional relevant uses of the word include the far advanced pregnancy of a female mammal or capital letters. Additionally there are a couple of informal and colloquial meanings plus some locally restricted and variety dependent usages, all of which beyond the scope of the paper at hand. Nevertheless it would be interesting to investigate on these remote senses of the lexical item *big* as well and to find out to what respect these aspects can be seen as synonymous to *large* and *great* and whether synonymy differs in local dialects and varieties of English. All in all the OED lists an astonishing 133 senses of the adjective *big*. Many of these, however, are obsolete or exceptionally rare. The OED hypothesizes about the etymology of the adjective and suggests that it is very likely an instance of borrowing. It is assumed that it is borrowed from early Scandinavian *bugga* which means 'mighty', 'rich' and 'wealthy'. Even though this explanation seems reasonable especially as it is consistent with the first attested meanings of the item *big*, the OED admits that this assumption is far from certain. Nevertheless it seems likely on semantic grounds and is not yet falsified.

As we can see, *big* is not only a word that can modify SIZE literally or metaphorically, it is also a polysemous word, even though most meanings ceased to exist through narrowing of meaning, and it will be interesting to see if the corpus-based study seconds these assumptions. As for the next item under investigation, *large*, the allegedly most synonymous concept of *big*, the dictionaries include the following information.

The Oxford English Dictionary defines *large* as “[o]f considerable or *relatively great* size, extent, or *capacity*” or “[o]f wide range or scope” (OOD). While the primary definition only differs in the addition of *relatively great* as qualification to the OOD entry on *big*, the secondary meaning presented is entirely different. The primary definition is still mainly based on physical size although a large capacity does not necessarily go hand in hand with actual size. For instance, a 1GB hard drive is considerably larger than a 512 MB hard drive, which is about half the memory size of the former, in terms of memory capacity yet it does not necessarily have to be physically larger than the latter. In spite of that, however, it is safe to assume that the first definition thoroughly covers physical size. The second definition to be found in the OED, “of wide range or scope”, is a very broad and wide definition which basically could comprise any type of metaphorical sense. It seems that the second definition of *large* is considerably broader than the one for *big*. A possible inference could be that due to the relative broadness of the category, *large* can be and is used more often than its close relative *big* in general due to the fact that it exhibits a wider array of possible applications. The sequence of synonyms reads *big*, *great*, *huge*, etc.

The Collins English Dictionary entry for *large* also features an entry based on physical size as primary meaning and mentions a wide or broad scope as alternative usage of the adjective. For the sake of completeness it has to be mentioned that the OOD also includes an obsolete meaning and a nautical application, both of which presumably not witnessed in the corpus data. If however this is the case, these instances will not be withheld. While the OOD and CED entries on *big* at least featured some differences, the entries on *large* are virtually identical.

The OED, owing to its comprehensive and historical character, again provides the most elaborate and therefore most fruitful entry. First of all, the OED lists 129 distinct senses for *large* as opposed to 133 for *big*. Starting from the premise that today's main meaning of both, *big* and *large*, is their modification of physical size, it could be said that *big* and *large* are absolute synonyms in this very respect. However, the corpus analysis reveals that even if *big* and *large* are used in their prototypical sense of modifying physical size, they differ in fine nuances of meaning and connotational aspects which means, given the fact that they are used in a context where their substitution does not change the truth value, they are, following Cruse's (1986) terminology, propositional synonyms. Moreover, if *big* exhibits 133 possible meanings and *large* 129, it is, diachronically speaking, impossible for the two items to be synonymous as they do not share all of their meanings at any given time. According to Murphy (2003), two or more words sharing a number but not all of their senses can be classified as sense synonyms. In either case, *big* and *large* do not exhibit absolute synonymy, absolute sameness of any meaning in every respect. This is already the first argument against the sameness of meaning of *big* and *large* even though the two items are mostly treated as synonymous in lexicography. Linguistics, cognitive linguistics that is, has already established that there has yet to be discovered a pair of absolutely synonymous words as meaning also always depends on the context and connotation. Autonomous linguists on the other hand, who claim that meaning exists independent from context and who do not consider connotation as exceedingly important, would probably still claim that there are words that are absolutely synonymous as these scholars follow a distinctly different paradigm.

According to the OED, *large* is derived from French *large*, meaning 'broad' and wide. The most important sense of the word is listed in the dictionary as:

In modern English, a general designation for considerable magnitude, used instead of great when it is not intended to convey the emotional implication now belonging to that word. (See great adj. 1) The more colloquial or less refined synonym is big. (OED)

This is a very interesting entry from a linguistic point of view as it puts forward a number of notions. First of all it is claimed that *great* conveys an emotional

implication which does not adhere to *large*. An interesting aspect for future studies on the concept of 'great' yet, while worth mentioning not too relevant for the paper at hand. The second assumption, however, is very relevant. If *big* really is the more colloquial and less refined synonym of *large*, a preponderance of the latter in academic and formally written texts is to be expected whereas in spoken texts presumably *big* is used more often. In summary, the OED entry on *large* features a number of now obsolete meanings of the concept such as 'generous' and 'munificent', is congruent with its emphasis on a wide range or capacity instead of the emphasis on physical size as in *big* and, interestingly enough, only lists *big* as synonym, *huge* cannot be found in the OED thesaurus.

The last item on the list, *huge*, is defined as "extremely large in size, amount or scope" (CED) by the Collins English Dictionary and as "extremely large; enormous" (OOD) in the Oxford Online Dictionary. According to these entries *huge* is a qualification of *large* or, in other words, *large* could be a hyperonym of *huge* and therefore the notion of absolute synonymy could be rejected. Additionally, *huge*, seems to have one meaning and one meaning only which could be indicative of *huge* having a good chance of being synonymous with *big* or *large* in certain respects. Furthermore it is absolutely noteworthy that *huge*, while being listed as synonymous to both, *big* and *large*, in the CED as well as the OOD, itself does not list the other two adjectives as synonyms.

The OED entry for *huge* is also characterized by its brevity. Even the OED only lists a single non-obsolete meaning for the adjective which is to be "very great, big, large or immense" for both, material and abstract things. Again, this entry could be interpreted in a way that proposes *huge* to be a qualification of *big* and in this case also *large*. Consequently *big* and *large* could be considered hyperonyms of *huge*. The two obsolete meanings mentioned in the OED prove that the item underwent a semantic narrowing although these dated meanings already were semantically very close to today's meaning of *huge*.

To sum up, while the OOD and the CED, monolingual dictionaries intended for a general audience which are thus mainly concerned with current word meanings, treat the three adjectives as relatively synonymous, even though they do list a

literal meaning of the word in modifying physical size and hint at slight differences in the adjectives' metaphorical modification of size. A look at the more refined and comprehensive OED already reveals that the adjectives are not as similar as the other dictionaries might suggest. First of all, while *huge* is being listed as synonymous to both, *big* and *large*, *huge* itself does not list *big* and *large* as synonymous which denies mutual interchangeability, a necessary property of absolute synonyms. Moreover, the OED lists a couple of current meanings for the items *big* and *large* whereas *huge* seems to have only a single non-obsolete meaning. This again hints at near-synonymy or propositional synonymy as actual synonyms would have to share any and every denotational and connotational meaning. Additionally, if for actual synonymy to be possible diachronic synonymy, analogous to sameness of language, would be regarded a crucial element, neither *big* nor *large* or *huge* could be considered synonymous as all three items exhibit genuine semantic changes and developed and dropped independent meanings. However, synonymy is customarily seen as a synchronic phenomenon.

As can already be seen from this lexicographical analysis of the synonymy of *big*, *large* and *huge*, more comprehensive dictionaries of English not exclusively aimed towards general users of English provide a more complex account of the three adjectives under examination and do not treat them as full synonyms. Nevertheless, for an even more comprehensive and all-encompassing account of the synonymy of the three adjectives another methodology had to be utilized, the corpus-based behavioral profile approach.

3.2 Behavioral profile analysis

The first corpus search that was carried out was what is arguably the most prototypical corpus query besides collocation analyses, the retrieval of raw frequencies of the items under examination and their normed occurrences. Table 6, very similar to the table provided by Biber et al. (1998: 44), lists the raw frequencies of the adjectival occurrences of *big*, *large* and *huge* within the Corpus of Contemporary American English as well as their occurrences divided by genres.

Table 6: Raw frequencies of *big*, *large* and *huge* in the Corpus of Contemporary American English

	Raw counts	Normed per million words
Total Sample		
(464,020,256 words)		
<i>big</i>	206,965	445.72
<i>large</i>	125,819	270.96
<i>huge</i>	49,677	106.98
Spoken		
(95,565,075 words)		
<i>big</i>	58,082	607.77
<i>large</i>	11,178	116.97
<i>huge</i>	12,557	131.40
Fiction		
(90,429,400 words)		
<i>big</i>	44,947	497.04
<i>large</i>	19,483	215.45
<i>huge</i>	10,865	120.15

	Raw counts	Normed per million words
<i>Magazine</i>		
(90,429,400 words)		
<i>big</i>	46,742	489.14
<i>large</i>	38,327	401.08
<i>huge</i>	11,515	120.50
<i>Newspaper</i>		
(91,717,452 words)		
<i>big</i>	49,044	534.73
<i>large</i>	23,375	254.86
<i>huge</i>	11,272	122.90
<i>Academic</i>		
(91,066,191 words)		
<i>big</i>	8,150	89.50
<i>large</i>	33,456	367.38
<i>huge</i>	3,468	38.08

Source: Author's illustration, based on Biber et al. (1998: 44)

A lot of insight on the behavioral profile of *big*, *large* and *huge* can already be gained from a simple frequency analysis. The search string “big.[j]” yields all occurrences of the word *big* as part of speech adjective. In the approximately 460-million-word Corpus of Contemporary American English *big* occurs 209,965 times which equals roughly 446 instances per million words. *Large* occurs 125,819 times which equals 271 instances per million words. Compared to Biber et al. (1998) who extracted 230 occurrences per million words for *big* and 408 for *large* from a 5.7-million-word sample from the Longman-Lancaster corpus it can be said that the COCA shows a higher number of words per million for the

adjective *big* than the sample extracted from the Longman-Lancaster corpus but, in turn, exhibits a smaller number of occurrences of the item *large* than the Longman-Lancaster corpus. Many reasons for this interesting, not necessarily predictable fact are possible. First of all, the Longman-Lancaster corpus is topic-driven rather than genre-driven like many of its counterparts including the COCA (Summers, 1991: 3-7). While this may be only a superficial difference in the corpus' structural organization and not influence the distributional patterns of types and tokens, one of the determining factors for the distinct frequency counts is the composition of the Longman-Lancaster corpus. It is constructed to be representative of the written discourse of the English language from the 1900s onwards and consequently does not comprise spoken texts. However, the approximately 607 tokens per million words for *big* found in the spoken section of the COCA have to be factored in when analyzing the entire sample as they considerably raise the overall count within the COCA. The same is true for the relatively low number of *large*-occurrences within said genre in the COCA which lowers the overall frequency and therefore partly accounts for the lower count in the corpus. Even more insight into the matter can be gained through close investigation of the distribution by genres.

As Biber et al. (1998) only distinguished between academic texts and fiction, the starting point for comparison are these two genres. The COCA lists approximately 90 occurrences per million words for *big* and 367 for *large* in its academic section as opposed to 31 for *big* and 605 for *large* in the respective section of the Longman-Lancaster corpus. Roughly 497 instances per million words of *big* and 215 for *large* are to be found categorized as fiction in the COCA compared to 408 and 232 in the Longman-Lancaster corpus. These numbers exhibit relative congruency and prove that within academic writing a preponderance of the adjective *large* can be attested and in fiction *big* is the more widely used variant. In other words, the results of the frequency analysis for *big* and *large* within the genres of academic writing and fiction in the COCA support the results brought forward by Biber et al (1998). However, the proposed higher frequency of *large* in the total sample of the Longman-Lancaster corpus cannot be verified due to the fact that, even if the entire genre of spoken language would have been factored out of the frequency analysis in the COCA as the Longman-

Lancaster corpus does not feature said genre, the COCA exhibits a higher number of tokens for the adjectives *big* than for the type *large*. For a more precise analysis of the reasons for this inversion of results a close examination of the representativeness of both corpora and the exact compilation of the genres and the selection of texts within the genres would be necessary.

As for the adjective *huge*, just as hypothesized, due to its monosemous nature and the fact that it is merely a qualification of its alleged synonyms *big* and *large*, its number of occurrences is relatively low throughout all genres featured in the COCA with the most occurrences per million words in spoken language and the least in academic texts.

These frequency distributions suggest that due to the preponderance of the word *big* in spoken discourse and its low number of usages in academic and therefore high register language, it inherently carries a certain collocational connotation. Additionally, the analysis at hand shows that *large* is used more often in an academic context than in spoken discourse in total numbers and also compared to its alleged synonymous counterpart *big*. It can already be deduced from this simple frequency analysis that the two items are not absolutely synonymous as suggested by the evidently genre-dependent distribution of the adjectives. Moreover, the notion put forward by the OED that *big* is the “more colloquial or less refined synonym” (OED) is also supported by the current findings. If, however, this notion holds true against a more elaborate and comprehensive behavioral profile analysis is discussed a few paragraphs below.

A chi-squared test for goodness-of-fit carried out using the language R environment employed in order to statistically prove what is immediately apparent from table 6 confirms that the adjectives are not evenly distributed throughout the genres. A language R test for independence indicates that the two variables, genre and type of adjective are associated and therefore a statistically relevant effect of the variable genre on the variable type of adjective can be seen which means that the distribution of the adjectives is not realized through mere chance or within the limits of a normal distribution. The variable genre has a relevant effect on the type of adjective employed which means that in general the

adjectives are not fully interchangeable and therefore not absolutely synonymous. The R commands for both calculations as well as the results can be found in the appendix.

The collocation analysis of *big*, *large* and *huge* yielded the following results, presented in tables 6, 7 and 8.

Table 7: Five most frequent collocates of *big*

Lexeme	Deal	Bang	Fan	Bucks	Leagues
Frequency of modification	6306	1680	1245	1091	782
Overall frequency	87610	5144	15651	8169	3756
Percentage	6.55	32.66	7.95	13.36	20.82
Mutual Information Score	3.03	5.35	3.31	4.06	4.71

Source: Author's Illustration (2014)

Table 8: Five most frequent collocates of *large*

Lexeme	Bowl	Heat	Numbers	Cup	Skillet
Frequency of modification	3793	3791	3569	2733	1626
Overall frequency	33471	52364	48804	57247	5631
Percentage	11.33	7.24	7.31	4.77	28.88
Mutual Information Score	4.54	3.69	3.91	3.29	5.89

Source: Author's Illustration (2014)

Table 9: Five most frequent collocates of *huge*

Lexeme	Amounts	Fan	Profits	Sums	Losses
Frequency of modification	590	444	243	221	220
Overall frequency	12430	15651	12283	2771	12902
Percentage	4.75	2.84	1.98	7.98	1.71
Mutual Information Score	4.62	3.88	3.36	5.37	3.15

Source: Author's Illustration (2014)

These tables show the five most frequent collocations for the adjectives *big*, *large* and *huge*. Usually collocation analyses and their respective outcomes are ordered by mutual information score as a high number of collocations can still turn out to be relatively irrelevant if the number of general, absolute occurrences of the lexeme is multiple times higher. In such a case, while in absolute numbers the lexeme is frequently modified by the item under investigation, due to its high circulation, the collocation of the two items is of no further interest. In this particular study however, ordering by mutual information score has proven to be inadequate as especially phrases and utterances that occur only a handful times in the corpus but are exclusively modified by either adjective exhibit a high mutual information score yet are not relevant for the study at hand. An example for this phenomenon is, for instance, the phrase *big ten*, which is an American Collegiate sports league. The lexeme *ten* only occurs a couple of times within the corpus yet, caused by the fixed expression, it is modified frequently by *big* which, in turn, results in a high mutual information score. Nevertheless, only collocations with a mutual information score higher than three have been considered so as to maintain statistical relevance. Moreover, the top collocation for *big* was, in fact, *big* and for *large*, *large*, a fact mainly caused by two reasons. First of all, the rhetorical device of repetition for emphasis is exploited. An example of this phenomenon is given in (9).

- (9) It's monster madness at the Kiel Auditorium. Big trucks, big, big trucks. Trucks so big other trucks can ride around inside of them. (COCA: Rapp, Adam: *Finer Noble Gases*. Fiction. Playscript, 2001)

Additionally, due to the fact that the collocation analysis has been carried out for a lexical window of 18 words, which is one of the many advantages of a corpus-based behavioral profile analysis over conventional methodologies, results based word repetition via deixis and anaphoric references had to be singled out and barred from the analysis. The outcomes of the collocation searches by these standards are presented in the tables above and, upon close inspection, provide additional information on the behavioral profiles of the adjectives.

First and foremost it is quite striking that *large* is prone to modify concrete, inanimate objects belonging to the semantic field of cooking such as *bowl*, *cup* and *skillet*. In these cases the type of modification is literal as opposed to metaphorical. As a matter of fact, even the noun *heat* which, even though not directly modified by *large* but frequently co-occurring with the adjective within the 18-word window, mostly occurs in the semantic context of cooking. Obviously some usages stem from other fields such as technology or environment but a good amount can be classified as belonging to the semantic field of cooking. The presence of the inanimate, abstract category of *numbers* in the collocational profile of *large* proves that it is not only used to modify concrete objects but is also capable of modifying abstract categories.

As for the collocations of *big* the analysis suggests that they are predominantly abstract in nature. Moreover, the collocation with *leagues* is presumably one that is only to be found within the data of the COCA as it is a Corpus of American English since the Americans label their premier hockey, baseball, basketball and football leagues as the big leagues and therefore a preponderance of the collocation especially in the genres newspaper and magazine comes as no surprise. On closer investigation also the spoken texts included in the COCA show a close collocation of *big* and *leagues* as they consist of recorded transcripts of unscripted conversations on TV and radio shows such as morning

shows that, by nature, contain discussions on professional sports and are therefore prone to contain the collocation.

In terms of lexical field, *huge* seems to share much of its behavioral profile with *big* which is remarkable as *big* and *large* are generally considered by linguists to be more synonymous yet they distinctly differ in their predominant lexical fields. First of all, both adjectives frequently co-occur with *fan*, a lexical relation that can be subsumed under the semantic field of sports. Secondly, both adjectives seem to be connected to the lexical field of finance as *huge* collocates with *profits* and *sums* and *big* collocates with *bucks*. The collocations *amounts* and *losses* for *huge* and *deal* for *big* respectively can also, while not being exclusively related to the lexical field of *finance*, at least partly be attributed to the very semantic field.

To sum up, it can be said that, astonishingly, from a semantic point of view *huge* shares more traits with *big* than *big* does with *large* which is absolutely remarkable. Moreover, judged from the five most frequent collocations, *large* seems to be the only item to customarily modify SIZE both, literally and metaphorically. It also seems to be more inclined to modify concrete concepts whereas *big* and *huge* generally modify abstract categories. Consequently, for their most frequent collocations, the analysis suggests that semantically *big* and *huge* are more synonymous than *big* and *large* or *huge* and *large*, a previously unheard of fact and a very relevant finding of the study at hand. The question whether this inference from collocational analyses holds true against an even more elaborate and all-encompassing method of investigation and can be reinforced is presented in what follows.

As suggested by the power analysis, which can be found in the appendix, a sample of 65 random occurrences, equally distributed within the genres featured in the COCA, has been extracted and manually annotated for morphological, semantic and syntactic properties as presented in table 3. These ID-tags have then been analyzed using the language R environment for statistical computing. Subsequently the data is presented graphically. Due to the fact that most behavioral profile approaches utilize hierarchical agglomerative cluster analyses and therefore there already exists a number of studies following said

methodology (cf. Gries and Otani, 2010), the paper at hand tries to break the mold and incorporates a method called multiple correspondence analyses. Multiple correspondence analyses require a certain type of data and therefore, to successfully conduct a multiple correspondence analysis and its according graphical representation, a number of steps have to be followed. First of all, as mentioned in the section on behavioral profile analyses, the manually annotated data has to be entered in a spreadsheet software as in table 9. The full table can be found in the appendix.

Table 10: *Excerpt of the ID-Tags for big, large and huge*

Lexeme	Countability	Negation	Size modification	Connotation	Semantic field
big	countable	not present	metaphorically	positive	professional
big	not countable	not present	metaphorically	positive	undisclosed
big	countable	not present	literally	neutral	leisure
large	not countable	not present	metaphorically	positive	professional
large	not countable	not present	metaphorically	positive	professional
large	countable	not present	metaphorically	positive	professional
large	not countable	not present	metaphorically	positive	leisure
large	not countable	not present	metaphorically	positive	financial

Source: Author's Illustration (2014)

Subsequently, the table has to be loaded into R and converted into a numerical cross-tabulation or pivot table, the format needed for R to convert numerical correlations into distances in a two-dimensional plot. An excerpt of the flat dataframe and its raw frequencies of the variants by variable is provided in table 10, the full dataframe can be found in the appendix.

Table 11: *Excerpt of the R dataframe*

Lexeme	Genre	Tense	Voice	Sentence type
big:65	acad:39	future:5	active:195	declarative: 191
huge:65	fiction:39	past:65		interrogative:4
large:65	mag:39	present:125		
	news:39			
	spoken:39			

Source: Author's Illustration (2014)

While the simple, flat dataframe does not provide information other than a general summary of the imported data, the transformation into a cross-tabulation already provides some insight into the behavioral profile of *big*, *large* and *huge*.

While neither morphological aspects such as voice, with all of the 195 examples being tagged as active, or tense, with all three adjectives being equally used mostly in the present and past tense, nor syntactic considerations such as sentence type, clause type or type of adjective show a significant difference between the items which therefore can be considered synonymous from morphosyntactic point of view, differences in the semantic structure are evident. An excerpt of the results of the cross-tabulation is presented in table 10.

Table 12: Excerpt of the *R* pivot table

Category	big	huge	large
Type of object: abstract	39	50	28
Type of object: concrete	26	15	37
Object: animate	9	5	14
Object: inanimate	56	60	51
Countability: countable	51	45	51
Countability: uncountable	14	20	14
Size modification: literal	22	20	48
Size modification: metaphorical	43	45	27
Connotation: positive	18	25	21
Connotation: neutral	32	10	35
Connotation: negative	15	30	9

Source: Author's Illustration (2014)

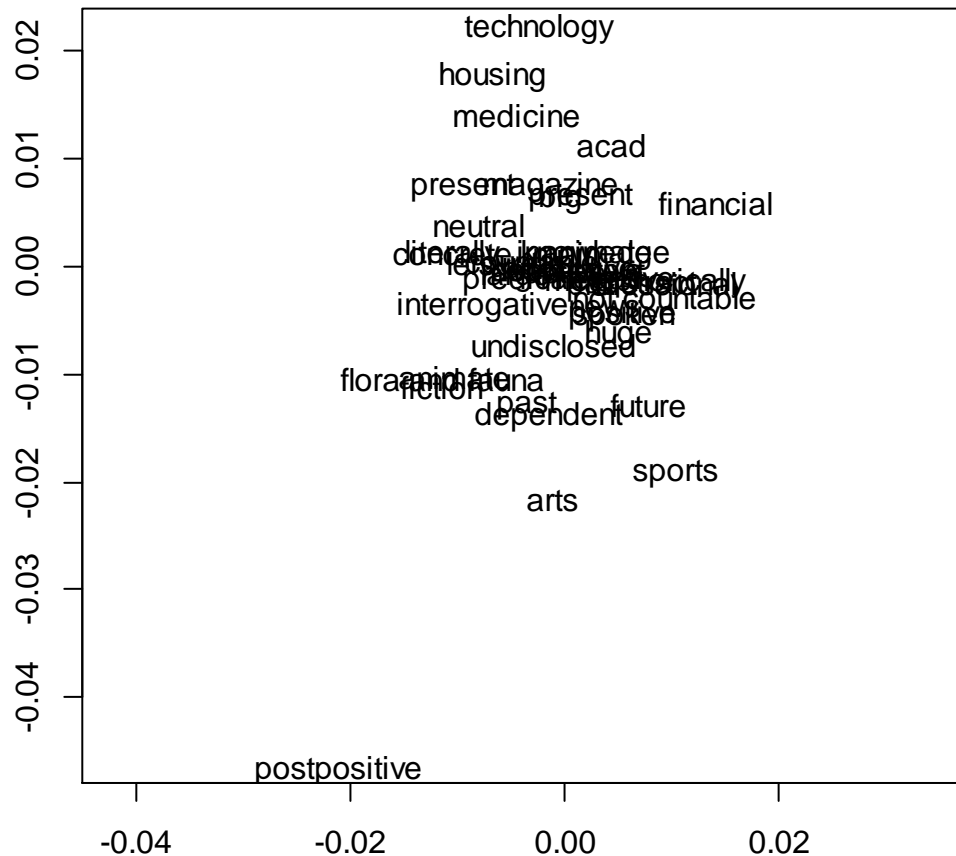
As mentioned above, while the variants of the variables constituting the morphosyntactical aspects are relatively evenly distributed among the adjectives and therefore not distinctly different behavioral profile for either of the items can be deduced, *big*, *large* and *huge* exhibit semantic differences.

While all three adjectives are capable of modifying concrete as well as abstract objects, the analysis shows that *big* and *huge* are more prone to modifying abstract objects whereas *large* shows a slight preference for concrete concepts as object for modification. Upon closer inspection the pivot table proves that within the sample in fiction *large* is used almost exclusively to modify concrete objects. As the sample is tailored to be representative of the general population it can be deduced that in general American English *large* in fiction is almost exclusively used to modify concrete objects as well which, in turn, means that from this, semantic aspect, *big* and *huge* are more synonymous than *big* and

large, the two synonymous considered to exhibit a higher overall level of synonymy according to previous studies. Moreover, while *big* and *huge* seem to favor metaphorical modification of SIZE, which can partly be attributed to their inclination of modifying abstract objects, *large* more frequently modifies SIZE literally. An example for a literal modification of an abstract concept is the utterance 'large crowd'. While a crowd is an abstract concept, a large crowd is actually an entity that physically occupies space and therefore SIZE is modified literally in such a case. Lastly, *big* and *large* are customarily used in all connotational contexts whereas *huge* tends to convey either a positive or negative implication and is only rarely used neutrally. Thus, it could be deduced that this fact is the reason why *huge* is in limited circulation in the academic genre as academic language is supposed to be neutral and clear of connotational implications, suggestive concepts and cultural assumptions.

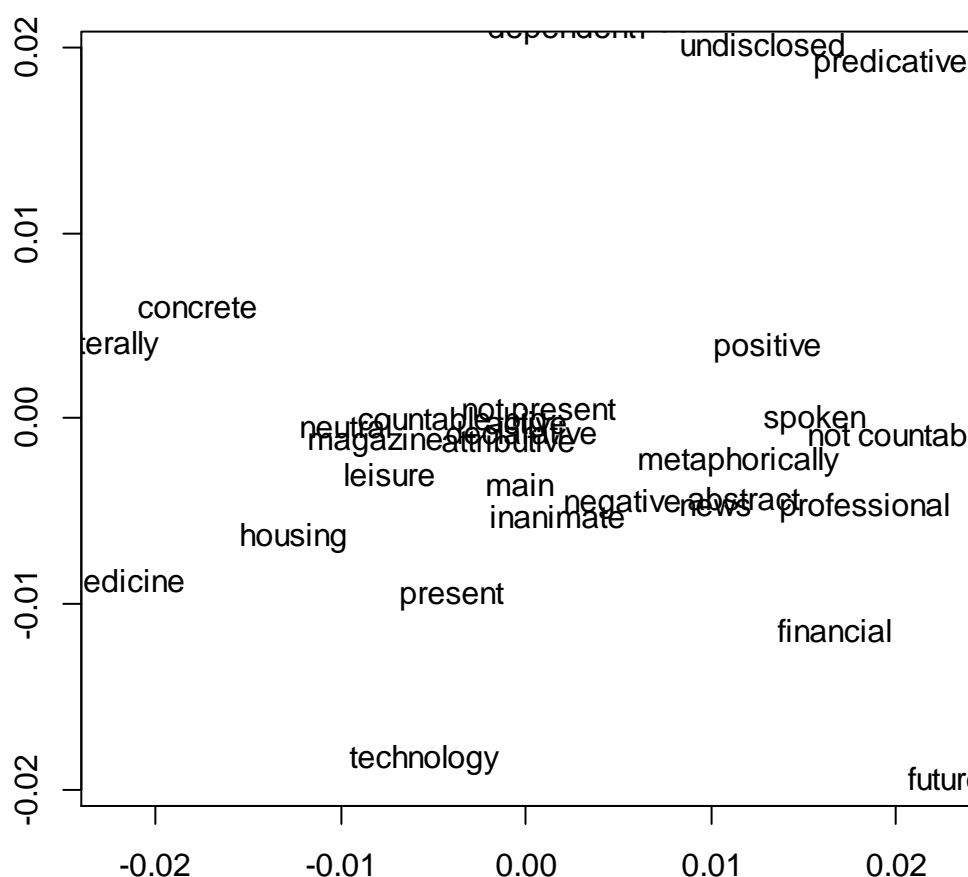
As the last step of the behavioral profile analysis of the adjectives *big*, *large* and *huge*, the cross-tabulation has been turned into multiple correspondence analysis for the full dataset and a number of subsets in order to compare selected aspects and to provide valid and interpretable graphic representations of the findings.

Figure 2: Multiple Correspondence Analysis full dataset



Source: Language R plot of the dataframe (2014)

The first analysis contains the full dataset. Unfortunately, due to the relative synonymy and therefore statistical similarity of the three concepts which is converted into a two-dimensional plot by the multiple correspondence analysis the plot is hardly interpretable. To overcome these difficulties, three autonomous behavioral profile analyses for each adjective have been carried out.

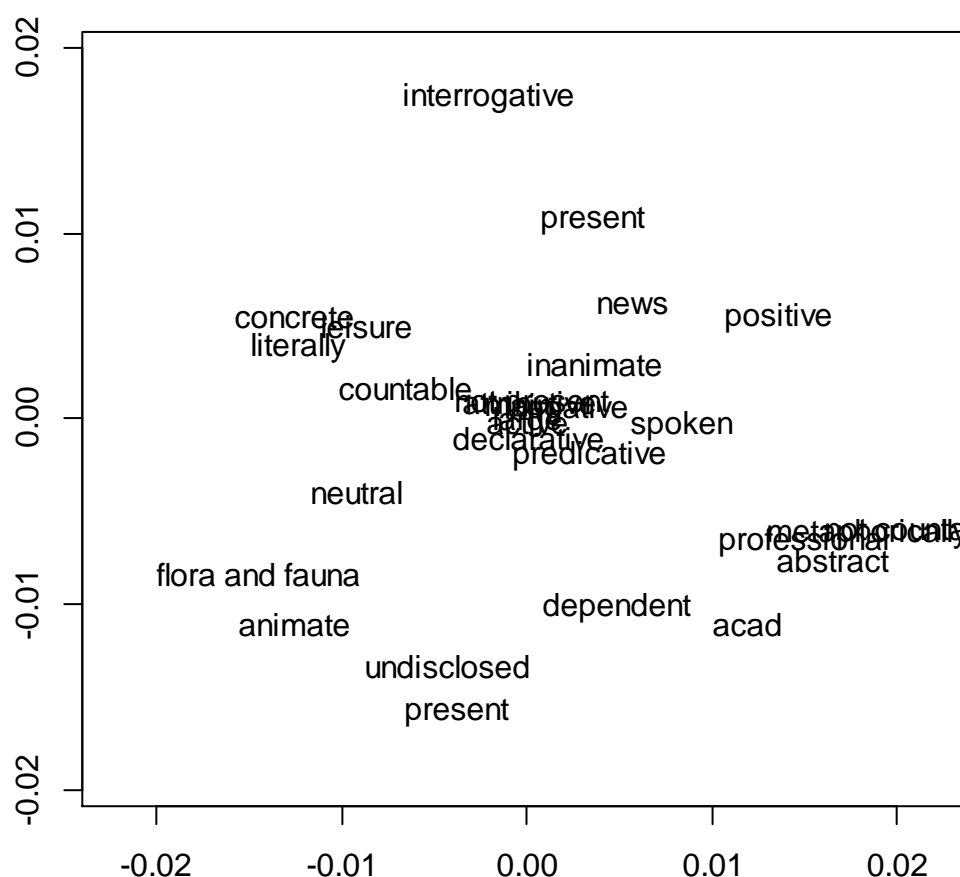
Figure 3: Multiple Correspondence Analysis BP big

Source: Language R plot of the subset *big* (2014)

The plot graphically depicts the results of the cross-tabulation and can therefore be seen as enhancing the pivot table. The adjective *big* is used in neutral contexts and to a slightly lower extent in negative or positive environment. The data shows that in 65 occurrences, *big* was used 32 times neutrally, 18 times positively and 15 times negatively which means approximately fifty percent of all usages of *big* are neutral in connotation. Moreover, *big* is slightly more prone to modify abstract concepts than concrete objects. Additionally, which is true for virtually any adjective in the English language, it is predominantly used attributively but hardly ever postpositively. Semantically *big* is used in many

different fields but, out of the ten semantic fields identified in the study (see appendix for full pivot table), shows an inclination towards the fields of leisure, professionalism and technology. Figure 4 depicts the respective graphical representation for the adjective *large*.

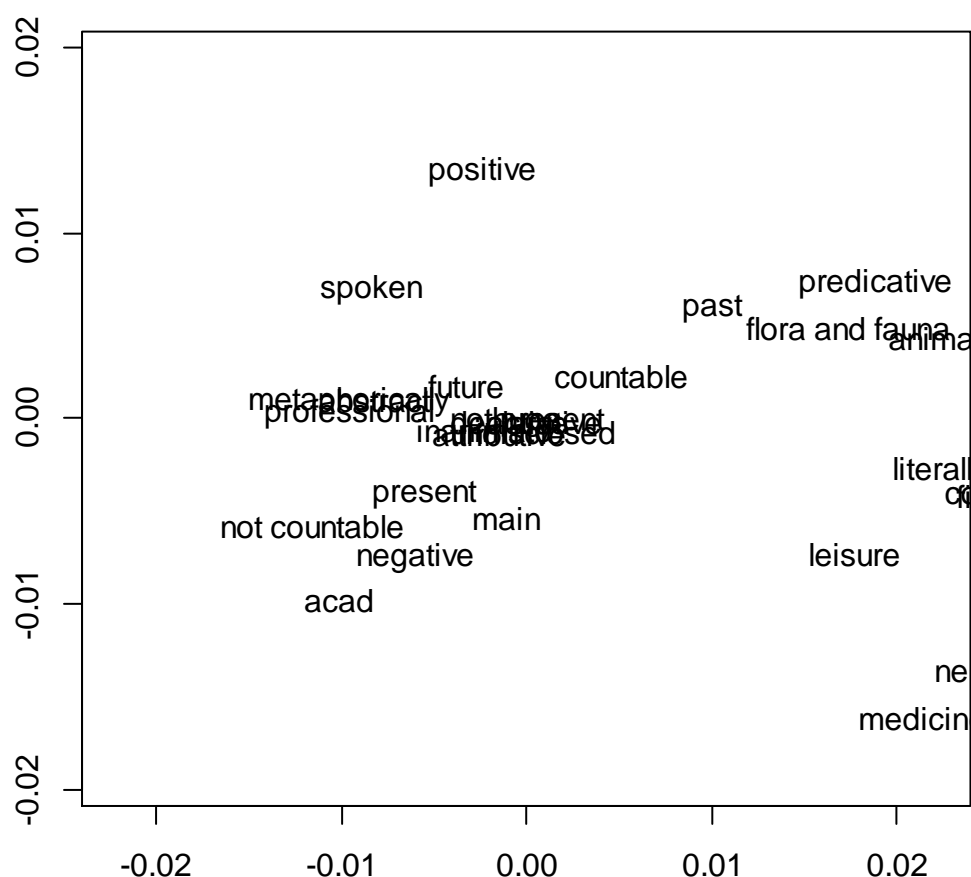
Figure 4: Multiple Correspondence Analysis BP *large*



Source: Language R plot of the subset *large* (2014)

The multiple correspondence analysis for the adjective *large* is, in fact, as expected and hypothesized, very similar to the one for *big*. As previously established on the basis of the cross-tabulation, contrary to *big*, *large* slightly prefers to modify concrete items rather than abstract concepts. Consequently

large, as opposed to *big*, often modifies SIZE literally. Both phenomena can be seen in the plot as the relative distance of the concept *abstract* to the center point of the figure, where, hardly recognizable, *large* is positioned, is slightly larger than the distance between *concrete* and the mid-point. Semantically *large* is often employed within the fields of leisure and professionalism, similar to *big*, but also within flora and fauna, a field not even to be found in the multiple correspondence analysis of *big* within reasonable boundaries. On the other hand however, *large* is rarely utilized within the semantic field of technology, contrary to its similar concept *big*. Figure 4 depicts the behavioral profile of *huge*, which is distinctively different from the other two items.

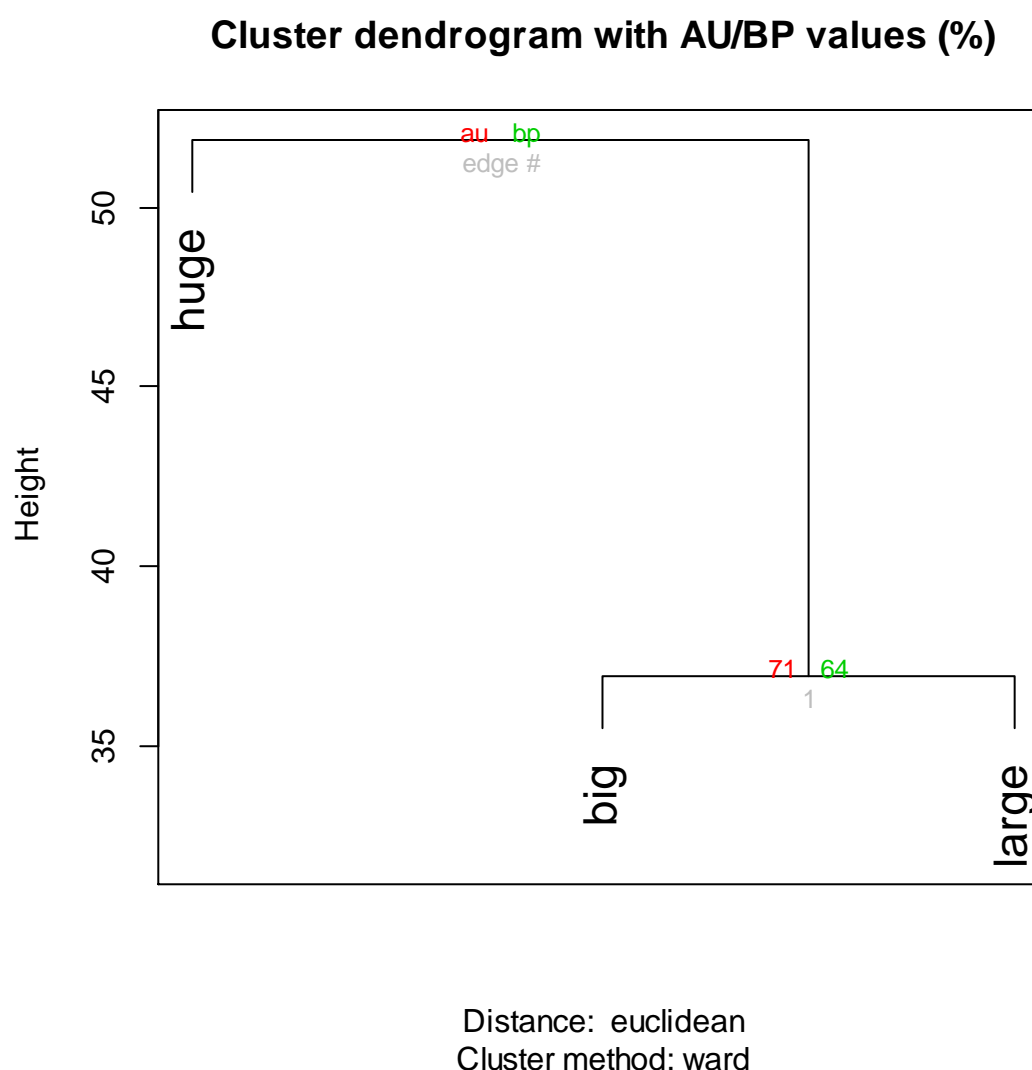
Figure 5: Multiple Correspondence Analysis BP huge

Source: Language R plot of the subset *huge* (2014)

Already at the first glance it can be seen that while, as established, even *big* and *large* do not exhibit an entirely synonymous behavioral profile, *huge* is considerably distinct. Morphosyntactically *huge* is still very similar to and can, even in a fine grained analysis be considered synonymous to *big* and *large*. However, *huge* is even more likely to metaphorically modify abstract categories than *big* and, again, distinctively different to *big* and *large*, is hardly used in a neutral context and almost always carries a positive or negative connotation which is presumably the reason why it scarcely occurs in academic texts, a high register text type which is supposed to be objective and free of connotational implications.

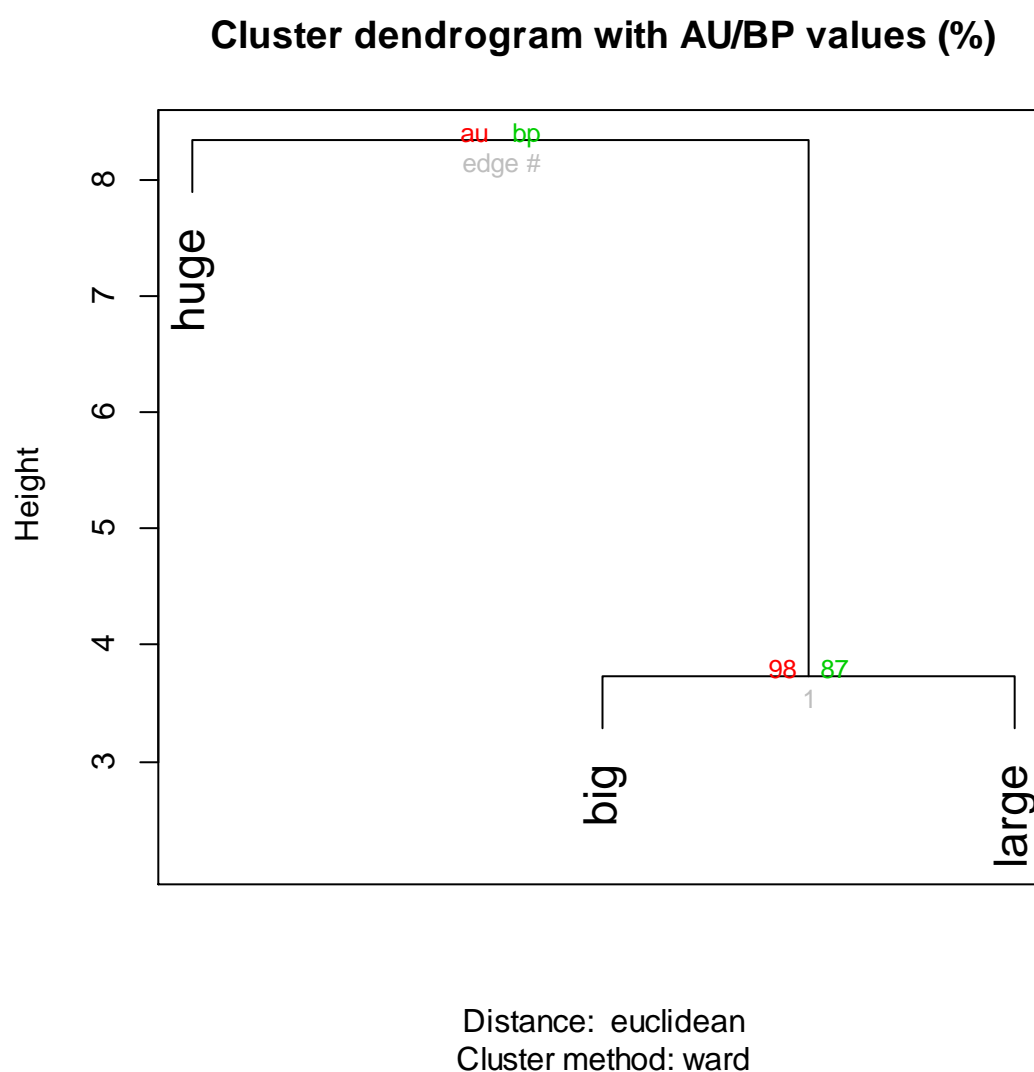
To sum up, the multiple correspondence analyses graphically depict and therefore facilitate the interpretation of the pivot tables. Recapitulating, it can be said that all three adjectives exhibit what can be considered as absolute synonymy in morphology and syntactic structure but distinctively differ in their semantic properties. Following the framework proposed by Cruse (2004), due to the fact that truth conditions are not changed when substituting one item with another, *big*, *large* and *huge* can be classified as absolute synonyms in a morphosyntactical respect and propositional synonyms in a semantic context. As cognitive linguists take the view that the meaning of a word is context dependent and not autonomous, the logical deduction is that, generally speaking, *big*, *large* and *huge* are not fully synonymous or near-synonyms but propositional synonyms, synonyms that are identical in their prototypical denotational meaning but exhibit differences in their connotational aspects as proven in the study. With respect to Murphy's (2003) taxonomy and the question of polysemy, this paper is not capable of classifying the three adjectives according to the proposed categories as the analysis only covered the prototypical denotation of the three adjectives, the modification of SIZE. Therefore a scientifically valid statement on whether the three lexical items are full synonyms, synonyms in every sense(s) of the concepts. However, as suggested by the consulted dictionaries, especially the OED, this is very unlikely. Nevertheless this study has proven that *big*, *large* and *huge* are relatively synonymous in their prototypical denotational meaning and can therefore be considered as sense synonyms according to Murphy's classification. In order to statistically verify the outcomes of the pivot tables and the multiple correspondence analysis, a cluster analysis has been conducted.

Figure 6: Cluster analysis of *big*, *large* and *huge*



Source: Language R dendrogram (2014)

The cluster analysis proves that *huge*, *big* and *large* are synonymous and additionally indicates that *big* and *large* are even more synonymous concepts. The AU and BP values, approximately unbiased and bootstrap probability value, indicate how strongly data supports the cluster on a scale from 1 to 100 (cf. Suzuki and Shimodaira, 2005). For overall synonymy, the score of *big* and *large* supports the outcomes of the paper and the dendrogram for the syntactic features, presented in figure 7, proves that syntactically *big* and *large* are even more synonymous exhibiting AU and BP values of 98 and 87 respectively.

Figure 7: Syntactic *cluster analysis of big, large and huge*

Source: Language R syntactic dendrogram (2014)

4 Discussion and theoretical implications

The assumption mentioned in the introduction, that actual synonyms are very rare if at all to be found, has clearly been reinforced by this study. Two of the canonically considered to be most synonymous lexical items, the adjectives *big* and *large*, and another very similar concept, *huge*, have been proven to be propositional synonymous. The behavioral profile approach utilized in the study at hand reinforces some notions provided by other, methodologically different studies, but also dismisses some previous findings. As for partial synonymy it can be said that the three adjectives exhibit synonymous traits from a syntactic point of view but are distinctively different from a semantic perspective and therefore cannot be considered absolutely synonymous in a holistic, lexico-semantic position. Generally, however, it can be seen that the behavioral profiles of the adjectives exhibit a large area of overlap and therefore constitute three very similar concepts. The paper at hand is another example of a successfully employed corpus-based behavioral profile analysis, a fairly recent methodology, which is capable of producing all-encompassing linguistic accounts of lexical items and therefore suitable for cognitive linguistic language analyses. The advent of electronic corpora allowed for previously unheard of methods and methodologies and, in connection with statistical computing, is a flourishing branch of linguistics. Corpora enable linguists to back up their qualitative data and analysis with quantitative considerations which, in turn, if done correctly and statistically appropriate, to extrapolate from the sample to the entire population. Thus corpus linguistics is a very powerful tool for any and every linguists.

Nevertheless, even corpus linguistics and consequently this study has its limitations. First of all, there is the question of representativeness of all kinds of corpora. The Corpus of Contemporary American English contains the category *spoken* which comprises unscripted, yet specific types of speech as they are transcribed from TV and radio programs and does therefore not fully resemble actual, spontaneous language use. Furthermore, corpora tend to present frequent occurrences as the most important ones when, in fact, very often the one, odd utterance is the interesting one worth exploring. Moreover, corpora can

hardly ever be up to date as it is just not possible for the linguists that run the corpora to add new data in monthly intervals, for instance.

Concerning this paper representativeness is also an issue. Even though the power analysis suggested 65 items per adjective as starting point for the analysis there is still a statistically relevant margin for error and, as always, larger sample sizes would mean even more substantial scientific evidence. Additionally, while morphosyntactic classification was a relatively straightforward process as these are somewhat clear-cut categories, semantic categorization proved to be more challenging and is by no means chiseled in stone and are open for discussion.

With these constraints in mind, however, the study at hand gives a good account of the behavioral profiles of *big*, *large* and *huge* and due to its fine grained analysis provides some insight into the semantic properties of the adjectives that was not provided by earlier and methodologically different studies, at least not in the way presented in this thesis. Nevertheless investigations on synonymy are far from finished and other studies, possibly utilizing other, even more sophisticated currently not available methodologies, will be conducted and, just like this study, provide insights that facilitate and improve the linguistic accounts on synonymy, near-synonymy and propositional synonymy.

5 References

A. Dictionaries and Corpora

Collins English Dictionary. (2010). Glasgow: Harper Collins. (CED, searched February 2014). Available at <http://www.collinsdictionary.com>

Concise Oxford Dictionary of Linguistics (2nd ed). (2007). Oxford: Oxford University Press. (CODL, searched February 2014). Available at <http://www.oxfordreference.com>

Davies, Mark (2008-). *The Corpus of Contemporary American English*. Available at <http://corpus.byu.edu/coca/>. (COCA; searched February 2014)

Oxford English Dictionary. (2006). Oxford: Oxford University Press. (OED, searched February 2014). Available at <http://oed.com>

Oxford Online Dictionary. (2014). Oxford: Oxford University Press. (OOD, searched February 2014). Available at <http://oxforddictionaries.com>

B. Other Literature

Abdi, Hervè.; Valentin, Dominique. (2007). "Multiple correspondence analysis". In Salkind, N. (Ed.) *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: SAGE Publications, 652-658.

Apresjan, Jurij D. (1973). "Synonymy and synonyms". In Kiefer, F. (ed.). *Trends in Soviet Theoretical Linguistics*. Dordrecht: Reidel, 173-200.

Bergenholtz, Henning; Gouws, Rufus. (2012). "Synonymy and synonyms in Lexicography". *Lexicographica*, 28(1): 309-336.

Biber et al. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Bolinger, Dwight L. (1968). "Entailment and the meaning of structures". *Glossa* 2, 119-127.

Butler, Christopher S. (1985). *Statistics in Linguistics*. Oxford: Basil Blackwell.

- Church, Kenneth et al.** (1994). "Lexical substitutability". In Atkins, B.T.S; Zampolli, A. (eds.). *Computational approaches to the lexicon*. Oxford and New York: Oxford University Press, 153-177.
- Church, Kenneth et al.** (1994). "Computational Approaches to the Lexicon". *Lexical substitutability*. Oxford: Oxford University Press, 153-177.
- Clark, Eve V.** (1992). "Conventionality and contrast: Pragmatic principles with lexical consequences." In: Kittay, E.F.; Lehrer, A. (eds.): *Frames, fields and contrasts: New essays in semantic and lexical organization*. Hillsdale NJ: 171-181.
- Cohen, Jacob.** (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edition). Routledge: London.
- Croft, William; Cruse, D. Alan.** (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Cruse, D. Alan.** (1986). *Lexical Semantics*. Cambridge University Press: Cambridge.
- Deese, James.** (1964). "The associative structure of some common English adjectives". *Journal of Verbal Learning and Verbal Behaviour* 1(2), 79-84.
- Dilin, Liu.** (2010). "Is it a chief, main, major, primary, or principal concern? A corpus-based behavioral profile study of the near-synonyms and its implications." *International Journal of Corpus Linguistics* 15, 56-87.
- DiMarco, Chrysanne; Hirst, Greame; Stede, Manfred.** (1993). *The semantic and stylistic differentiation of synonyms and near-synonyms*. University of Toronto: AAI Spring Symposium.
- Divjak, Dagmar S.; Gries, Stephan Th.** (2006). "Ways of trying in Russian: Clustering behavioral profiles". *Corpus Linguistics and Linguistic Theory* 2(1), 23-60
- Divjak, Dagmar S.; Gries, Stephan Th.** (2008). "Clusters in the mind? Converging evidence from near-synonym in Russian". *The Mental Lexicon* 3(2), 188-213

- Divjak, Dagmar S.; Gries, Stephan Th.** (2009). "Corpus-based cognitive semantics: A contrastive study of phrasal verbs in English and Russian. In K. Dziwirek and B. Lewandowska-Tomaszczyk (Eds.). *Studies in cognitive corpus linguistics*. Frankfurt am Main: Peter Lang, 273-296
- Edmonds, Philip; Hirst, Graeme.** (2002). "Near-synonymy and lexical choice". *Computational Linguistics* 28(2), 105-144.
- Firth, John.** (1957). *Papers in linguistics*. Oxford: Oxford University Press.
- Glynn, Dylan** (2010). "Corpus-driven cognitive semantics. An introduction to the field". In Glynn, Dylan; Fischer, Karen (eds.). *Corpus-driven cognitive semantics. Quantitative approaches*. Berlin: Mouton de Gruyter, 1-42.
- Glynn, Dylan; Robinson, Justyna.** (2012). *Corpus Methods in Cognitive Semantics*. Amsterdam: John Benjamins.
- Gries, Stefan.** (2010). "Useful statistics for corpus linguistics". In Sánchez, Alquilio; Almela, Moises (eds.). *A mosaic of corpus linguistics: selected approaches*. Frankfurt am Main: Peter Lang, 269-291.
- Gries, Th. Stefan; Otani, Naoki.** (2010). "Behavioral profiles: A corpus-based perspective on synonymy and antonymy". *ICAME Journal* 34, 121-150.
- Hanks, Patrick.** 1996." Contextual dependency and lexical sets". *International Journal of Corpus Linguistics* 1(1), 75-98.
- Harris, Zellig S.** (1970). *Papers in structural and transformational linguistics*. Dordrecht: Reidel.
- Kennedy, Graeme.** (1987). *Semantic interpretation and the resolution of ambiguity*. Cambridge: Cambridge University Press.
- Kishner, Jeffrey; Gibbs, Raymond.** (1996). "How just gets its meanings: polysemy and context in psychological semantics". *Language and Speech* 39(1), 19-36.
- Lakoff, George.** (1987). *Women, Fire and Dangerous Things. What categories reveal about the mind*. Chicago: University of Chicago Press.
- Leech, Geoffrey.** (1981). *Semantics*. (Second edition). Harmondsworth: Penguin.
- Mc Enery, Tony; Wilson, Andrew.** (2001). *Corpus linguistics: An introduction*. (2nd edition). Edinburgh: Edinburgh University Press.

-
- McCarthy, Michael; O’Keeffe, Anne.** (2010). “Historical perspective: What are corpora and how have they evolved?”. *Routledge Handbook of Corpus Linguistics*. New York: Routledge.
- Miller, George; Charles, Walter.** (1991). “Contextual correlates of semantic similarity”. *Language and Cognitive Processes* , 6(1), 1-28.
- Murphy, M. Lynne.** (2003). *Semantic Relations and the Lexicon*. Cambridge: Cambridge University Press.
- Murphy, M. Lynne.** (2013). “What we talk about when we talk about synonyms”. *International Journal of Lexicography* 26(3), 279-304.
- Sparck Jones, Karen.** (1986). *Synonymy and Semantic Classification*. Edinburgh: Edinburgh University Press.
- Storjohann, Petra.** (2009). “Plesionomy: A case of synonymy or contrast?”. *Journal of Pragmatics* 41(11), 2140-2158.
- Summers, Delia.** (1991). *Longman/Lancaster English Language Corpus: Criteria and Design*. Harlow: Longman.
- Suzuki, Ryota. Shimodaira, Hidetoshi.** (2006). “Pvclust: an R package for assessing the uncertainty in hierarchical clustering”. *Bioinformatics* 22(12), 1540-1542
- Tyler, Andrea; Evans, Vyvyan.** (2001). “Reconsidering Prepositional Polysemy Networks: The Case of over”. *Language* 77(4), 724-765.
- Winchester, Simon.** (2003). *The Meaning of Everything: The Story of the Oxford English Dictionary*. Oxford: Oxford University Press.

6 Appendices

Appendix 1: Lexicographical sources

RESOURCE (abbreviation)	ACCESSED VIA	MODALITY
Collins English Dictionary (CED)	collinsdictionary.com	electronic, online, free
Oxford Online Dictionary (OOD)	oxforddictionaries.com	electronic, online, free
Concise Oxford Dictionary of Linguistics (CODL)	oxfordreference.com	electronic, online, subscription accessed via UNIVE VPN
Oxford English Dictionary (OED)	oed.com	electronic, online, subscription accessed via UNIVIE VPN
Collins COBUILD Advanced British English Learner's Dictionary (COBUILD)	collinsdictionary.com learners	electronic, online, free

Appendix 2: Language R statistical tests

Language R chi-squared test for goodness-of-fit *big*

```
> types=c(58082,44947,46742,49044,8150)
> hypothetical=c(0.20,0.20,0.20,0.20,0.20)
> chisq.test(types,p=hypothetical)
```

Chi-squared test for given probabilities

data: types
X-squared = 35836.98, df = 4, p-value < 2.2e-16

Language R chi-squared test for goodness-of-fit *large*

```
> types=c(58082,44947,38327,49044,33456)
> hypothetical=c(0.20,0.20,0.20,0.20,0.20)
> chisq.test(types,p=hypothetical)
```

Chi-squared test for given probabilities

data: types
X-squared = 8153.156, df = 4, p-value < 2.2e-16

Language R chi-squared test for goodness-of-fit *huge*

```
> types=c(12557,10865,11515,11272,8150)
> hypothetical=c(0.20,0.20,0.20,0.20,0.20)
> chisq.test(types,p=hypothetical)
```

Chi-squared test for given probabilities

data: types
X-squared = 995.42, df = 4, p-value < 2.2e-16

Language R chi-squared test for independence

```
> spoken=c(58082,11178,12557)
> fiction=c(44947,19483,10865)
> magazine=c(46742,38327,11515)
> newspaper=c(49044,23375,11272)
> academic=c(8150,33456,3468)
> table=matrix(c(spoken,fiction,magazine,newspaper,academic),3)
> table
  [,1] [,2] [,3] [,4] [,5]
[1,] 58082 44947 46742 49044 8150
[2,] 11178 19483 38327 23375 33456
[3,] 12557 10865 11515 11272 3468
> chisq.test(table)
```

Pearson's Chi-squared test

data: table

X-squared = 53703.28, df = 8, p-value < 2.2e-16

Appendix 3: Language R pivot table full dataset

	genre.acad	genre.fiction	genre.magazine	genre.news	genre.spoken
big	13	13	13	13	13
huge	13	13	13	13	13
large	13	13	13	13	13
	tense.future	tense.past	tense.present	voice.active	
big	1	20	44	65	
huge	4	23	39	66	
large	0	22	45	67	
	sentence.type.declarative	sentence.type.interrogative			
big	64	1			
huge	66	0			
large	64	3			
	clause.type.dependent	clause.type.main	type.of.adjective.attributive		
big	9	56	62		
huge	6	60	61		
large	6	61	60		
	type.of.adjective.postpositive	type.of.adjective.predicative			
big	0	3			
huge	0	5			
large	1	6			
	type.of.object.abstract	type.of.object.concrete	object.animate		
big	39	26	9		
huge	50	16	6		
large	29	38	15		
	object.inanimate	countability.countable	countability.not.countable		
big	56	51	14		
huge	60	46	20		
large	52	51	16		
	negation.not.present	negation.present	size.modification.literally		
big	64	1	22		
huge	66	0	20		
large	63	4	40		

size.modification.metaphorically connotation.negative connotation.neutral			
big	43	15	32
huge	46	30	10
large	27	10	35
connotation.positive semantic.field.arts semantic.field.financial			
big	18	2	1
huge	26	0	2
large	22	0	1
semantic.field.flora.and.fauna semantic.field.housing			
big	1	1	
huge	7	0	
large	10	1	
semantic.field.knowledge semantic.field.leisure semantic.field.medicine			
big	3	12	7
huge	2	8	3
large	1	15	1
semantic.field.professional semantic.field.sports			
big	10	2	
huge	24	6	
large	20	1	
semantic.field.technology semantic.field.undisclosed			
big	15	11	
huge	2	12	
large	4	13	

Appendix 4: Dataframe part 1 – morphosyntactic analysis

lexeme	genre	tense	voice	sentence type	clause type	type of adjective
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	acad	present	active	declarative	main	attributive
big	fiction	past	active	declarative	dependent	attributive
big	fiction	present	active	declarative	main	attributive
big	fiction	present	active	declarative	dependent	attributive
big	fiction	past	active	declarative	dependent	attributive
big	fiction	past	active	declarative	dependent	attributive
big	fiction	past	active	declarative	main	attributive
big	fiction	past	active	declarative	main	attributive
big	fiction	present	active	declarative	main	predicative
big	fiction	present	active	declarative	main	attributive
big	fiction	past	active	declarative	main	attributive
big	fiction	present	active	declarative	main	attributive
big	fiction	past	active	declarative	main	attributive
big	fiction	past	active	interrogative	main	attributive
big	magazine	present	active	declarative	dependent	attributive
big	magazine	past	active	declarative	main	attributive
big	magazine	present	active	declarative	main	attributive
big	magazine	present	active	declarative	main	attributive
big	magazine	present	active	declarative	main	attributive
big	magazine	present	active	declarative	main	attributive
big	magazine	past	active	declarative	main	attributive
big	magazine	past	active	declarative	dependent	attributive
big	magazine	present	active	declarative	main	attributive
big	magazine	past	active	declarative	main	attributive
big	magazine	present	active	declarative	main	attributive
big	magazine	present	active	declarative	main	attributive
big	magazine	present	active	declarative	dependent	attributive
big	news	past	active	declarative	dependent	attributive

big	news	present	active	declarative	main	attributive
big	news	present	active	declarative	main	attributive
big	news	past	active	declarative	dependent	attributive
big	news	present	active	declarative	main	attributive
big	news	present	active	declarative	main	attributive
big	news	past	active	declarative	main	attributive
big	news	present	active	declarative	main	attributive
big	news	present	active	declarative	main	attributive
big	news	present	active	declarative	main	attributive
big	news	future	active	declarative	main	attributive
big	news	present	active	declarative	main	attributive
big	news	present	active	declarative	main	attributive
big	spoken	past	active	declarative	main	attributive
big	spoken	past	active	declarative	main	attributive
big	spoken	past	active	declarative	main	attributive
big	spoken	present	active	declarative	main	attributive
big	spoken	past	active	declarative	main	attributive
big	spoken	present	active	declarative	main	attributive
big	spoken	present	active	declarative	main	attributive
big	spoken	present	active	declarative	main	attributive
big	spoken	present	active	declarative	main	predicative
big	spoken	present	active	declarative	main	attributive
big	spoken	present	active	declarative	main	attributive
big	spoken	past	active	declarative	main	predicative
big	spoken	present	active	declarative	main	attributive
large	spoken	present	active	declarative	dependent	predicative
large	spoken	present	active	declarative	main	predicative
large	spoken	present	active	declarative	main	attributive
large	spoken	present	active	interrogative	main	attributive
large	spoken	present	active	declarative	main	attributive
large	spoken	present	active	declarative	main	attributive
large	spoken	present	active	interrogative	main	predicative
large	spoken	present	active	declarative	main	attributive
large	spoken	present	active	declarative	main	attributive
large	spoken	present	active	declarative	main	attributive
large	spoken	past	active	declarative	main	attributive
large	spoken	past	active	declarative	main	attributive
large	fiction	past	active	declarative	main	attributive
large	fiction	past	active	declarative	main	postpositive
large	fiction	past	active	declarative	main	attributive
large	fiction	past	active	declarative	dependent	attributive
large	fiction	past	active	declarative	main	attributive
large	fiction	past	active	declarative	main	attributive
large	fiction	past	active	declarative	main	attributive
large	fiction	past	active	declarative	main	attributive

[illegible]

huge	fiction	present	active	declarative	main	attributive
huge	fiction	past	active	declarative	dependent	predicative
huge	fiction	past	active	declarative	main	attributive
huge	fiction	past	active	declarative	main	attributive
huge	fiction	past	active	declarative	main	attributive
huge	spoken	present	active	declarative	main	attributive
huge	spoken	present	active	declarative	main	attributive
huge	spoken	present	active	declarative	main	attributive
huge	spoken	present	active	declarative	main	attributive
huge	spoken	present	active	declarative	main	attributive
huge	spoken	present	active	declarative	main	attributive
huge	spoken	present	active	declarative	main	attributive
huge	spoken	future	active	declarative	main	attributive
huge	spoken	past	active	declarative	main	attributive
huge	spoken	past	active	declarative	main	attributive
huge	spoken	past	active	declarative	main	attributive
huge	spoken	past	active	declarative	main	attributive

Appendix 5: Dataframe part 2 – semantic analysis

type of object	object	countability	negation	size modification	connotation
concrete	inanimate	countable	present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable not	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	animate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	negative
concrete	inanimate	countable	not present	literally	negative
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral

abstract	inanimate	countable not	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	metaphorically	negative
concrete	animate	countable not	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	animate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	animate	countable	not present	literally	negative
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	positive
abstract	inanimate	countable	not present	metaphorically	negative
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable not	not present	literally	neutral
abstract	inanimate	countable not	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	animate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable not	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative
concrete	inanimate	countable	not present	literally	positive
abstract	inanimate	countable not	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable not	not present	metaphorically	positive
concrete	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	neutral
concrete	animate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable not	not present	metaphorically	negative
concrete	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	neutral
abstract	animate	countable not	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	animate	not	not present	metaphorically	positive

		countable			
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative
		not			
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	neutral
		not			
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	animate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	negative
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
		not			
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	animate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	positive
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	neutral
concrete	animate	countable	present	literally	negative
concrete	inanimate	countable	present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
abstract	animate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	positive

concrete	animate	countable	not present	literally	positive
abstract	inanimate	countable	present	metaphorically	neutral
concrete	inanimate	countable	not present	literally	positive
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	neutral
		not			
concrete	inanimate	countable	not present	literally	negative
		not			
abstract	inanimate	countable	present	metaphorically	neutral
abstract	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	negative
		not			
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	positive
		not			
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	negative
concrete	inanimate	countable	not present	literally	positive
concrete	animate	countable	not present	metaphorically	neutral
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	negative
		not			
abstract	inanimate	countable	not present	metaphorically	neutral
		not			
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	inanimate	countable	not present	literally	negative
		not			
abstract	inanimate	countable	not present	metaphorically	negative
		not			
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	literally	neutral
		not			
abstract	inanimate	countable	not present	metaphorically	positive
abstract	animate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	neutral
concrete	animate	countable	not present	literally	neutral

abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable not	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	negative
concrete	inanimate	countable	not present	literally	positive
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	literally	negative
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable not	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable not	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative
concrete	inanimate	countable	not present	literally	negative
abstract	inanimate	countable not	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable not	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	neutral
abstract	inanimate	countable not	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	negative
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	positive
abstract	animate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	literally	positive
abstract	inanimate	countable	not present	metaphorically	negative
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable not	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable not	not present	metaphorically	negative
abstract	inanimate	countable not	not present	metaphorically	negative
abstract	inanimate	countable not	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative

abstract	animate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
concrete	animate	countable	not present	literally	negative
concrete	inanimate	countable	not present	literally	neutral
concrete	inanimate	countable	not present	literally	neutral
abstract	inanimate	countable	not present	metaphorically	neutral
		not			
abstract	inanimate	countable	not present	metaphorically	positive
		not			
concrete	animate	countable	not present	literally	negative
abstract	animate	countable	not present	literally	positive
concrete	inanimate	countable	not present	literally	positive
abstract	inanimate	countable	not present	metaphorically	negative
concrete	animate	countable	not present	literally	neutral
		not			
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	negative
		not			
abstract	inanimate	countable	not present	metaphorically	negative
concrete	inanimate	countable	not present	literally	positive
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	negative
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	metaphorically	neutral
		not			
abstract	inanimate	countable	not present	metaphorically	positive
concrete	inanimate	countable	not present	literally	negative
abstract	inanimate	countable	not present	metaphorically	positive
		not			
abstract	inanimate	countable	not present	metaphorically	negative
		not			
abstract	inanimate	countable	not present	metaphorically	positive
abstract	inanimate	countable	not present	literally	negative

7 Abstract in German

Diese Arbeit beschäftigt sich mit der Synonymie dreier häufig gebrauchter englischer Adjektive aus dem semantischen Feld der Größe (SIZE), *big*, *large* und *huge*. Diese drei Begriffe werden von vielen Wörterbüchern und Thesauri als synonym angesehen. Das Ziel dieser Arbeit ist es einen vollständigen Überblick über die Eigenschaften der Adjektive aus sowohl semantischer als auch syntaktischer Perspektive zu liefern. Zu diesem Zweck wird die Methodik der Behavioral Profile Analysis angewandt.

Um ein vollständiges Verhaltensprofil zu erstellen wurde die statistische Verteilung der Adjektive im Corpus of Contemporary American English analysiert, Konkordanz sowie Kollokation erfasst, analysiert und interpretiert. Eine statistisch relevante Anzahl an Beispielen wurde manuell kommentiert und mit Hilfe mathematischer und bildgebender Methoden ausgewertet. Die R Oberfläche für statistische Berechnungen bildet dabei die Basis und wurde verwendet um die statistischen Tests auszuführen, die Daten zu verarbeiten und die Ergebnisse graphisch darzustellen. Diese Ergebnisse wurden in weiterer Folge mit vorhergehenden Studien verglichen und kommentiert.

Die Studie zeigt, dass selbst die Adjektive *big* und *large*, zwei Konzepte die weitestgehend als so synonym wie möglich angesehen werden, nicht vollständig synonym sondern maximal Nah-Synonyme oder Plesionyme sind. Nichtsdestotrotz gibt es Bereiche, in denen sich die Adjektive synonym verhalten als in anderen. Desweiteren ist zu erwähnen dass, obwohl sich *huge* einige Attribute mit *big* und *large* teilt, es sich im Gesamtprofil doch deutlich von den anderen beiden unterscheidet. Die Studie bestärkt die Annahme, dass kognitive Synonymie, Bedeutungsgleichheit in jedem einzelnen Aspekt, in natürlichen Sprachen unmöglich zu sein scheint. Gleichzeitig zeigt sie jedoch, dass teilweise Synonymie, Bedeutungsgleichheit bezogen auf bestimmte Aspekte, möglich ist, speziell unter bestimmten linguistischen Paradigmen wie Generative Linguistik oder Autonome Linguistik. Synonymie ist keine scharf

abgegrenzte linguistische Kategorie sondern ein Kontinuum das viele feine Bedeutungsnuancen enthält.

Curriculum Vitae

Personal Information

Name: Christian Moser, BA
Address: 1180 Vienna, Austria
Telephone: +43(0)650 9055559
Date & Place of Birth: 7 July, 1983, Spittal/Drau
Nationality: Austria

Education

Sept. 2012-present **MA in English Language and Linguistics**
at the University of Vienna
⑩ Department of English and American Studies
Language, Literature and Culture (15 ECTS); Advanced Academic
Language Skills (10 ECTS); Research Foundations (15 ECTS);
Applied Linguistics & TEFL (35 ECTS); Research Module (10 ECTS);
MA Thesis (30 ECTS); Defensio of Thesis (5 ECTS)

Title of qualification awarded: Master Degree (MA)

Sept. 2007- Sept. 2012 **BA in English and American Studies**
at the University of Vienna
⑩ Department of English and American Studies

Title awarded: Bachelor of Arts (BA)

1998-2003 **Upper Secondary School with an emphasis on natural sciences**
(BORG Spittal an der Drau)

1994-1998 **Lower Secondary School** with an emphasis on languages
(BG Spittal an der Drau)

1990-1994 **Primary School,**
9872 Millstatt am See

Relevant Work Experience

Oct. 2008-present

English as a foreign language teacher
at Dr. Rampitsch Language Institute

Foreign language skills & competence levels

English (Active & passive skills: C2))

Italian (Active skills: B1; passive skills: B2)

French (Active skills: A2; passive skills: B1)