

Arttu Luhtala

“YOU'RE SO ST... YOU'RE SO DUMB.”
A Corpus-linguistic Comparison of *dumb* and *stupid* in
American English

TIIVISTELMÄ

Arttu Luhtala: "You're so st... You're so dumb.": A Corpus-linguistic Comparison of *dumb* and *stupid* in American English
Kandidaatintutkielma
Tampereen yliopisto
Kielten kandidaattiohjelma, englannin opintosuunta
Toukokuu 2021

Tutkimuksen tarkoituksena oli englanninkielisten lähisynonymien *dumb* ja *stupid* adjektiivimuotojen käytön ja merkitysten eroavaisuuden ja samankaltaisuuden tutkiminen. Tutkimusaihe valittiin erityisesti kahdesta syystä. Ensiksi, vaikka sanat ovat usein määritelty toistensa synonyymeiksi sanakirjoissa, joissakin niistä edelleen painotetaan sanan *dumb* vanhahtavaa ja loukkaavaa puhekyvyttömyyteen viittaavaa merkitystä. Toisaalta kyseisen sanaparin käyttöä ei ole aiemmin tutkittu korpusaineiston avulla, ja loukkaavien sanojen käyttöä on ylipäänsäkin tutkittu korpuksen avulla suhteellisen vähän.

Tutkimuskysymykset liittyivät kolmeen aihealueeseen. Ensimmäisenä tutkittiin adjektiivien yleisyyttä amerikanenglannissa ja maailmanenglannissa sekä tekstilajien välillä amerikanenglannissa. Lisäksi tutkittiin sanojen kollokationaalisia piirteitä semanttisen prosodian ja preferenssin valossa. Kollokaatio viittaa sanojen tavanomaiseen yhteisesiintymiseen ja semanttinen prosodia kollokaatioissa piilevään, kirjaimellisen merkityksen ylittävään merkityksen ulottuvuuteen, kun taas semanttinen preferenssi niiden yleisten kollokaattien ryhmiin, joilla on yhteisiä piirteitä. Adjektiivien määritelmät myös selvitettiin neljästä verkkosanakirjasta ja kahdesta kielioppaasta.

Tutkimusaineistot kerättiin kolmesta englanninkielisestä korpuksista. Kahta korpuksista, uutisenglantiin keskittyvää Corpus of News on the Webiä ja yleistä internetenglantia sisältävää Corpus of Global Web-Based Englishiä käytettiin vertailemaan Yhdysvaltain englantia ja sen ulkopuolisia variantteja. The Corpus of Contemporary American English toimi pääasiallisena tutkimusaineiston lähteenä, josta haettiin adjektiivien kollokaatit, esiintymistaajuudet eri tekstilajeissa sekä kollokaattiesiintymien kontekstia täsmentävät konkordanssirivit.

Tutkimustulosten perusteella on pääteltävissä, että *dumb* ja *stupid* ovat hyvin lähekkäisiä synonyymejä kollokaattiansa ja käyttökontekstiensa samankaltaisuuteen viitaten. Sanojen välillä on silti erojakini, sillä *dumb* osoittautui heikommaksi loukkaavaksi adjektiiviksi kuin *stupid*. Tähän liittyen havaittiin, että sanan *dumb* kanssa vahvemmin kollokoivat sanat olivat yleensä merkitykseltään voimakkaampia kuin sanan *stupid* samanlaisen ryhmän kollokaatit. Sanan *dumb* amerikanenglantilaisuuteen viittasi sen merkittävimmät ainutlaatuiset kollokaatit, jotka olivat sanakirjojen mukaan amerikkalaisperäisiä. Lisäksi tuotiin esille se, että sana *dumbass*, jota käytetään samankaltaisesti kuin kollokaatiota *dumb ass*, yleistyi korpusaineistossa vuosina 1990-2019.

Avainsanat: korpuslingvistiikka, amerikanenglanti, loukkaukset, lähisynonymia, kollokaatio, semantiikka

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

CONTENTS

1	Introduction	1
2	Background material	2
2.1	Corpus linguistics	2
2.2	Adjectives	4
2.3	Semantic concepts	4
2.4	Dictionary entries.....	6
2.5	Usage guide entries.....	7
3	Methods	7
3.1	Corpora used.....	7
3.2	Data gathering	8
3.3	Data analysis	9
4	Findings and discussion	9
4.1	Frequency.....	10
4.2	Collocation	13
4.3	The case of <i>dumb</i> and <i>stupid ass</i>	18
4.4	The significance of “I may be dumb, but I’m not stupid.”	21
5	Conclusion	23
	Works Cited	27
	Appendix 1.....	29
Table 1	Per million frequencies of <i>dumb</i> and <i>stupid</i> in NOW and GloWbE in descending order by size of representation in NOW.....	10
Table 2	The per-genre frequencies of <i>dumb</i> and <i>stupid</i> in COCA.	11
Table 3	ADJ + NOUN collocate lists for <i>dumb</i> and <i>stupid</i> , ordered alphabetically.....	14

1 Introduction

Dumb and *stupid* are defined very similarly in dictionaries. In most, if not all of them, one of the definitions for *dumb* is *stupid*, which in turn is often defined as lacking intelligence or importance. However, *dumb* also means to ‘not be able to speak’, mainly referring to animals incapable of human speech, or mute people. The latter of these definitions is commonly thought to be offensive and dated, which is often mentioned along the definitions. These points will be further detailed in section 2.4.

Other near synonymic word pairs and sets have been studied by using corpus data for at least five decades. These words have generally been neutral in meaning, such as *almost* and *nearly* in Kjellmer (2003), *high* and *tall* in Taylor (2003), and *chief*, *major*, etc. in Liu (2010). Some have also examined words which differ in polarity, such as Partington’s (2004) investigation into *happen* and *set in*, where *happen* is found to be more neutral than *set in*. Contrast this to *dumb* and *stupid*, which both exhibit negative meaning. There seems to be very little to no research done, at least using corpus linguistics, on these kinds of insulting words. This may pose a problem, especially for learners of English when interacting with native speakers. The difference between “He’s a dumb student” and “He’s a mute student” is quite obvious to a skilled speaker, even though by dictionary definition, their literal meanings could be the same.

In this study, through corpus linguistics and semantic theory, the aim is to investigate how the words function in American English. Dictionary and study guide entries will provide background information on the uses of the adjectives. In addition, the frequency of the words will be compared to other variants of English to give further context on the usage of the adjectives in American English.

The study aims to answer three research questions. First, what are the frequencies of *stupid* and *dumb* like between American English and World Englishes, as well as genres in American English? Second, what kind of collocational overlap or variation can one see between *dumb* and *stupid*? And finally, what kinds of semantic (dis)similarities do *dumb* and *stupid* exhibit?

The following section will detail the background material used in constructing the thesis. The section after will explain the methods, including which corpora were used and why, and what kinds of choices were made in investigating the words. The fourth section will reveal the results of the analysis and discusses those results. Finally, the thesis will be concluded by answering the research questions and looking at future possibilities.

2 Background material

This section will present brief overviews on the major themes the present study is based on. These themes include corpus linguistics, adjectives, and semantic theory. In addition, four dictionary entries and two usage guide entries were consulted to provide context for the empirical analysis part of the study.

2.1 Corpus linguistics

Corpus linguistics is a methodology and not an approach to linguistics (Meyer 2002, xi; Gries 2009, 1225). While the method is not without flaws, it has generally been deemed to offer data-driven insights into naturally generated language (Gries 2009; Kara 2017, 97). Thus, corpus linguistics offers a promising platform for an initial examination of words such as *dumb* and *stupid*, which are by nature rich in meanings outside of the literal.

Corpus linguistics understandably depends on the existence of corpora. Simply put, a corpus is a systematically arranged collection of naturally constructed texts, i.e., by people

without prompting them to create that text for a corpus, that have been spoken or written (Gries, 2009; Weisser, 2016). There are many types of corpora. The following explanation follows the categories in Weisser (2016, chap. 2): diachronic or synchronic corpora; written, spoken, or mixed corpora; general or specific corpora; and static or dynamic corpora. Many of these categories contain subcategories, but those are not relevant to the current study. Diachronic corpora refer to historical corpora, and synchronic corpora to contemporary ones. Spoken and written corpora contain spoken and written texts, and mixed corpora combine the two. General corpora aim to include language samples from as many domains as possible, while specific corpora focus on a specific subset of language, like the News on the Web corpus focuses on the language of news on the internet. Static corpora are sometimes called snapshot corpora and are designed to remain unchanged and fixed in size. In contrast, dynamic, or monitor corpora, are designed to continuously add material as time passes, effectively monitoring the constantly evolving lexicon and grammar of, for example, the English language.

One of the central concepts in corpus linguistics is collocation. Collocation is the habitual co-occurrence of words (Weisser 2016, 212; Kara 2017, 98). Habitual here primarily refers to chance of co-occurrence beyond random. Collocations are typically examined between zero to three places to either side of the node word. For example, in ‘dumb student’ *student* is in the +1 spot to the right of the node *dumb*, whereas in ‘the student is so dumb’ it is in the -3 spot. Concordance lines and frequencies are additional analytical tools in corpus linguistics. A concordance line is the surrounding context of a node word. Frequencies are often analysed as both raw and normalised frequencies. In most large corpora the frequencies are normalised per one million words. Based on those frequencies and the size of the corpus itself, different statistical scores may be assigned. Probably the most widely used is the Mutual Information (MI) score. MI-score compares observed frequency to expected

frequency. It measures collocational strength, especially of idiomatic expressions. The common value which MI-score must reach to be considered significant is 3 (Weisser 2016, chap. 10.8.1; Kara 2017, 99).

2.2 Adjectives

Adjectives are essential in describing objects, concepts, and people, especially in written language. This is because the reader does not see what is described and must rely on information given to them or their own imagination. The kinds of information adjectives convey are diverse: descriptor adjectives inform the reader on the colour, size, weight, and other prototypical details of the described item, whereas classifier adjectives restrict what the item can be (Biber et al. 1999, 508). Additionally, adjectives, like words from other classes, can carry evaluative or attitudinal meanings. These are commonly divided into negative, neutral, positive and meanings (Partington 2004, 131). To exemplify this, compare ‘a fat cat’, ‘an obese cat’ and ‘a plump cat’. Whereas *fat* is quite distinctly negative, *obese* can be seen as more neutral without much evaluation. *Plump*, on the other hand, is quite positively affective, and could be used to refer to one’s beloved but plump pet cat, for example.

2.3 Semantic concepts

The definition of meaning in semantics is not an unchallenged one. According to Murphy, there are three types of semantic meaning: denotative or ‘literal’ meaning, connotative meaning, and social meaning (Murphy 2010, 32-33). Following this definition, denotative meaning is literal in the sense that it relates a lexical item to the concepts it refers to. This type of meaning is close to the ones commonly seen in dictionaries. On the other hand, according to Murphy (2010, 33), connotative meaning is built on top of denotative meaning: it loosely associates one concept with another and is not really a part of denotative meaning.

Lastly, social meaning relates utterances and people, as it connects the speaker and the situation to that utterance. Denotative and connotative meanings are the primary point of inspection in this thesis, as social meaning is difficult to ascertain only through corpora.

Synonymy is, at its core, the phenomenon of two words having the same meaning (Kreidler 2013, 65; Murphy 2010, 110). If only denotative meaning were to be considered, this absolute synonymy would likely be more common. However, this paints a too black-and-white picture of language. Taking connotative as well as social meaning into account, which both contribute to distinct meanings beyond the denotational of a particular word, it rarely is the case that two words have the exact same meaning on all levels (Kreidler 2013, 66-67; Murphy 2010, 110). Additionally, a concept seen as common to most if not all natural languages is the resistance to identical expressions (Murphy 2010, 112). Apart from absolute synonyms, there are so called ‘sense’ synonyms, which are identified as two words sharing at least one sense of meaning. Near-synonyms, then, are words where some of these senses overlap and they may be substitutable in many or most cases, but not all (Murphy 2010, 110-111). *Dumb* and *stupid* certainly share at least one sense, that of foolishness, but do not share at least the sense of ‘incapability of human speech’.

Semantic prosody and preference both affect the meaning of an utterance. The distinction between an utterance and a word is made here, because these two concepts involve the surroundings of words. Semantic prosody is commonly defined as the primary carrier of evaluative and attitudinal meanings (Xiao and McEnery 2006, 106). According to Xiao and McEnery (2006), semantic prosody has a ‘staining’ effect. Returning to the example of *cat*, the negatively affective *fat* and positively affective *plump* certainly influence the otherwise neutral *cat*. Semantic preference, on the other hand, can be seen as the groups or categories of frequent collocates of a word with a common meaning (Partington 2004, 145). Partington continues by exemplifying the concept through the intensifying adjective *sheer*.

Five groups were identified, one of which was “‘magnitude’, ‘weight’ or ‘volume’” and another “expressing ‘force’, ‘strength’ or ‘energy’” (2004, 145).

2.4 Dictionary entries

Four dictionaries were used to gather comprehensive information on the adjectival definitions of *dumb* and *stupid*. Four online dictionaries were used: the *Lexico.com* English dictionary, the *Merriam-Webster.com Dictionary*, the *Cambridge Advanced Learner’s Dictionary & Thesaurus*, and the *Collins COBUILD Advanced English Dictionary*. Hereinafter, the dictionaries will be referred to as *Lexico*, *Merriam-Webster*, *Cambridge*, and *Collins*.

The dictionaries mostly agreed on the definition of *dumb*. The senses common in all dictionaries were, with varying wording and in no particular order: ‘stupid’, ‘foolish’ or ‘lacking intelligence’; someone ‘unable to speak’; and a ‘temporary inability or unwillingness to speak due to shock or surprise’. *Lexico*, *Merriam-Webster* and *Cambridge* also included the sense for *dumb devices*, which “cannot process data on their own”. All dictionaries classified the sense referring to a mute person as offensive and, excluding *Collins*, dated. On the other hand, unlike *Cambridge* and *Lexico*, *Merriam-Webster* did not classify the sense ‘stupid’ as informal nor mainly present in North America or the United States. *Collins* varies here as well, in that the sense ‘stupid’ is classified as informal but not markedly North American, but its unique sense of something silly and annoying is both. Additionally, *Collins* and *Merriam-Webster* include the sense of expressing something without words. Finally, *Merriam-Webster* provides the senses ‘silent’ and ‘lacking some usual attribute or accompaniment’.

The dictionary entries for *stupid* were simpler, although they varied more in nuances. The two senses all four dictionaries shared were ‘lacking intelligence, common sense or good judgement’ and ‘annoying or boring’. The latter sense was accompanied in *Merriam-Webster*

and *Collins* by ‘uninteresting’ and in *Cambridge* by ‘causing a problem’. In addition, *Lexico* provided the sense ‘dazed and unable to think clearly’ and *Merriam-Webster* that of ‘dulled in feeling or sensation’.

2.5 Usage guide entries

In addition to dictionary entries, some of which included usage notes alongside the definition, entries of *dumb* and *stupid* were also examined in two usage guides: the fourth edition of *Garner’s Modern English Usage* and the third edition of *Fowler’s Concise Dictionary of Modern English Usage*. This was done to gain further insight into the way *dumb* and *stupid* should be used from the viewpoint of a native speaker. While both guides include entries for *dumb*, only *Garner’s* includes one for *stupid*. Both mention that *dumb* is much more strongly associated with stupidity than its original meaning of speech-impairment (Garner 2016, 311; Butterfield 2016). The entry on *stupid* in *Garner’s* refers to the entry on *ignorant*. The entry makes the distinction between the two that “*stupid* refers to a lack of innate ability, whereas *ignorant* refers merely to a lack of knowledge” (Garner 2016, 482).

3 Methods

This section will detail the corpora used, provide reasons and aims why those corpora were used, and finally how the data used in the thesis was gathered and analysed.

3.1 Corpora used

Data was gathered from three corpora: the Corpus of News on the Web (NOW) (Davies 2016-), the Corpus of Global Web-Based English (GloWbE) (Davies 2013), and the Corpus of Contemporary American English (COCA) (Davies 2008-). NOW and GloWbE are some of the largest commonly available corpora available, as they contain more than 12.4 billion

and 1.9 billion words, respectively. COCA is one of the largest commonly available general corpora at around one billion words. As of April 27, 2021, the online interface for COCA (<https://www.english-corpora.org/coca/>) states that it is “the only large, genre-balanced corpus of American English.”

The NOW and GloWbE corpora were only used to investigate general frequencies across their twenty varieties of English, including American English. COCA, on the other hand, provided most of the collocational data and concordance lines, along with genre-specific frequencies. The primary reason why the frequencies from NOW and GloWbE were included in the analysis was to see whether *dumb* shows more prominence in American English than in other varieties.

3.2 Data gathering

The data from NOW and GloWbE was gathered in a simple manner. Both adjectives were queried in the online interface at www.english-corpora.org using the Chart function. The results, both raw and normalised per-country frequencies were then saved in a Google Sheets sheet. The frequency data was limited to countries which are represented by more than 300 million words in the NOW corpus, which resulted in data from thirteen countries. This was done to exclude English variants with very small representations, as they are not the highlight of this study. The resulting frequency data of the thirteen countries is revealed in section 4.1.

COCA was the main source of data for the thesis. Most of the analysis was done on a collocate list, procured by querying both *dumb* and *stupid* in the Collocates tool. Only noun collocates one word to the right of the node were examined. The collocate lists were constrained to the 100 most frequent collocates. All collocates in both lists still retained a Mutual Information score above three, which is often regarded as the minimum value for relevant collocations as detailed in section 2.1. All concordance lines, which were used to

examine the differences in collocation and meaning between *dumb* and *stupid*, were chosen randomly using a random number generator, specifically the one found by searching “random number generator” in Google on the Chrome browser. In addition to these collocate lists, queries were made for two phrases: “I may be dumb ,” and “I may be stupid ,”. The results and importance of these queries will be further discussed in section 4.4. All gathered data was saved and analysed in a Google Sheets file.

3.3 Data analysis

The collocate lists were first split into unique and common collocates. The collocates were then divided into five categories: animals, people, insults, concretes, and abstracts. Some collocates were seen to possibly be included in two or more categories, especially people and insults. Lastly, collocations were identified as strongly favoured or not favoured. This was done by first determining the differences in MI-score of common collocates. These differences were then turned into absolute values. Finally, the median value of the absolute values was calculated, which was found to be 0.475. In order to determine which node word a collocate prefers, the median value was doubled to 0.95. This means that if a collocate prefers *dumb*, the MI-score of *dumb* and the collocate is at least 0.95 higher than the MI-score of *stupid* and that collocate. This method was chosen, because the MI-scores were already provided by the COCA interface and they facilitate the comparison of this study and future research.

4 Findings and discussion

This section details the findings of the analysis and discusses them. These discussions are divided in four parts: the frequencies of *dumb* and *stupid* across variants of English and

genres in COCA; the collocational behaviour of the adjectives; the odd case of *dumb ass*; and the significance of the phrase “I may be dumb, but I’m not stupid.”

4.1 Frequency

Based on the data in Table 1 below, it does seem that some countries outside of the United States do resist the use of *dumb*, especially in news language. In more general language from GloWbE, both words are used more frequently in the US than in the other countries. This may correspond to the information on the words’ uses as presented in the dictionary entries, where it was explained that the sense of ‘annoying’ is mostly seen in this region. However, this study does not investigate this possibility in more detail. Future studies more focused on comparing different variants of English might clarify this further.

Table 1. Normalised frequencies (per million words) of *dumb* and *stupid* in NOW and GloWbE in descending order by size of representation in NOW.

REGION	NOW		GLOWBE	
	<i>dumb</i>	<i>stupid</i>	<i>dumb</i>	<i>stupid</i>
UNITED STATES	6.97	13.48	28.71	91.85
GREAT BRITAIN	4.1	16.76	9.93	54.93
CANADA	4.6	11.07	12.34	38.48
INDIA	2.51	5.87	10	28.91
IRELAND	2.47	12.51	6.87	34.83
AUSTRALIA	8	17.91	16.29	56.1
SOUTH AFRICA	3.25	13.49	7.85	36.13
NIGERIA	3.55	11.73	15.87	45.73
SINGAPORE	2.54	7.41	13.47	47.94
NEW ZEALAND	7.16	18.24	12.1	41.34
PHILIPPINES	2.53	8.58	11.12	37.9
PAKISTAN	3.16	9.97	10.2	35.68
MALAYSIA	4.53	18.84	12.58	59.43

As the data in Table 1 shows, *stupid* is much more common, often many times more so, than *dumb* across these thirteen variants of English. The United States data constitutes a baseline, where there are two to three instances of *stupid* against one of *dumb*. Most countries fall somewhere around this ratio. Some particularly *dumb*-resistant regions seem to be Ireland, Great Britain, Malaysia, and South Africa. In Ireland, *dumb* is used approximately one time

to the five times of *stupid* in both corpora. In the news language of British English, this ratio is one to four and in general internet language, around one to five. The ratio is roughly one to four in South Africa and Malaysia in both corpora.

Of course, ratios are not the only important consideration. Comparing the normalised frequencies with each other, the countries particularly resistant to *dumb* mentioned previously certainly continue the trend, as each of them exhibits a frequency of under 10 *dumbs* in one million words. Notably, the United States dominantly leads the charts in the GloWbE data of both words but is trumped in the NOW data. This is another avenue of future research to see whether this difference is purely linguistic or if there are extralinguistic factors affecting it as well. On the other hand, India seems to be resistant to both *dumb* and *stupid*, and a similar argument may be made here as above.

Moving on to COCA, more differences between the words were found. It was made clear that *stupid* is much more common generally than *dumb*, but now the focus shifts to the kinds of differences between genres. Both *dumb* and *stupid* are classified as informal, at least in certain contexts, so it should not be too surprising to see the results in Table 2.

Table 2. The per-genre frequencies of *dumb* and *stupid* in COCA.

Genre	<i>dumb</i>			<i>stupid</i>		
	FREQ	WORDS (MIL)	PER MIL	FREQ	WORDS (MIL)	PER MIL
Blog	3908	128.6	30.39	13399	128.6	104.18
Web	3161	124.3	25.44	9721	124.3	78.24
TV/Movie	5503	128.1	42.97	24626	128.1	192.28
Spoken	1515	126.1	12.01	3568	126.1	28.29
Fiction	3327	118.3	28.12	10294	118.3	87
Magazine	1294	126.1	10.26	2496	126.1	19.8
News	947	121.7	7.78	2026	121.7	16.64
Academic	320	119.8	2.67	580	119.8	4.84
All	19975	993	20.12	66710	993	67.18

One of the larger discrepancies here is the difference between the sections TV/Movies and Spoken. One possible reason for this is that the texts in Spoken are mainly transcripts of TV and radio talk shows, as is noted in the COCA Help section (<https://www.english->

corpora.org/coca/help/spoken.asp) as of April 27, 2021. As such, knowing that they were being recorded publicly, the word and topic choices in those programs may have varied from fully natural speech.

Another large difference is the frequencies in TV/Movies between *dumb* and *stupid*. Querying both adjectives and the wildcard * in the List function, it is revealed that both adjectives are used similarly, and especially often predicatively. The predicative, sentence-final use of *stupid* is particularly high, with 6,208 instances found. However, compared to *stupid*, *dumb* is encountered far less at 950 instances. The percentages of the sentence-final use out of the total number of instances are 25% and 17%, respectively. Furthermore, compare the concordance lines below:

- (1) And I thought running away and being with Jenna would fix everything, which of course that was dumb. (COCA, TV, 2016, *Awkward*.)
- (2) Yeah, people are dumb. (COCA, TV, 2017, *Gotham*)
- (3) You're so st... You're so dumb. (COCA, TV, 2018, *Goliath*)
- (4) I'm not stupid. (COCA, TV, 2019, *Legacies*)
- (5) One, this was stupid. (COCA, TV, 2019, *The Boys*)
- (6) Don't be stupid. (COCA, TV, 2019, *Tales of the City*)

The way the two adjectives are primarily used in these sentences are almost identical, so much so that the meaning between the sentence would hardly change if *dumb* was substituted for *stupid* or vice versa. Of particular note, however, is (3), where *stupid* has been repaired to *dumb*. This hints at a possible difference in intensity between the adjectives. Section 4.4 details this difference in more detail. What must be kept in mind here, though, is the effect of context and social meaning. These two variables may steer the speaker's choice of word in these situations. However, as the full context is difficult to discern from corpus data alone, it is not wise to conclude anything decisively from just these findings.

Comparing the data on American English news language from Table 1 and Table 2, there are small increases in frequencies in the COCA data. This may be partly explained by the inclusion of different time periods between the corpora. Whereas NOW only includes

data from 2010 onward, COCA provides data from 1990 until 2019. This could be one part of a larger historical look into the use of *dumb* and *stupid* in future research.

In academic texts, *stupid* and *dumb* are mostly used in quotations with the meaning of ‘stupid’. As for *dumb*, most of the other cases refer to its historical meaning, ‘mute’, although there are quite few of these overall. *Dumb luck* seems to be a fixed phrase, although not common at six instances, present in academic prose. The third largest contributor to this list, *dumb jock*, is in truth found twenty-six times in only four texts, where twenty-three of these are from a single text. Overall, the lack of *dumb* and *stupid* in academic texts is not against expectations, as academic writers tend to not overtly announce their opinions. On the other hand, many writers have probably identified the offensive and dated *dumb* in the ‘mute’ sense and have opted for other words.

4.2 Collocation

As seen previously, *dumb* and *stupid* exhibit similar meanings and functions, at least predicatively. This section aims to investigate their attributive use through a large list of collocates for both words retrieved from COCA. As mentioned earlier, the list consists of 100 nouns immediately to the right of the adjectives. This way, the analysis focuses on the kinds of nouns the adjectives modify. The list was also divided into thirty-six unique and sixty-four common collocates for both *dumb* and *stupid*. These lists are compiled into Table 3 below.

The full information retrieved from COCA for the thirty-five most frequent of both adjectives has been provided in Appendix 1. The common collocates contain many that refer to general concepts, like *thing(s)*, *stuff*, and *idea*. In fact, in terms of raw frequency, *things* is the highest ranking collocate for *stupid*, and third for *dumb*, only surpassed by *ass* and *luck*, of which *luck* is unique to *dumb*. While *ass* is not an uncommon collocate for *stupid* either, its usage is sometimes anomalous with *dumb*. This will be explored more in section 4.3.

Table 3. ADJ + NOUN collocate lists for *dumb* and *stupid*, ordered alphabetically.

Group	<i>dumb</i>	<i>stupid</i>
<i>unique</i>	asses, baby, beast, beasts, blonde, bombs, brute, bunny, butt, chick, cluck, cop, dogs, friends, fucks, fun, guy, guys, hick, jock, jocks, luck, masses, motherfuckers, nigger, ol (<i>ol'</i> , mistagged form of <i>old</i>), ox, pipe, redneck, shits, show, slut, stunt, terminal, terminals, waiter	accident, argument, bet, boat, book, brother, car, cat, choices, crap, dance, dream, face, fight, fool, hat, horse, idiots, job, laws, letter, machine, man, mouth, pet, plan, prank, remarks, rules, song, stories, video, wars, whore, woman, word
<i>common</i>	Americans, animal, animals, article, ass, asshole, bastard, bastards, bird, bitch, boy, choice, comment, comments, cow, cunt, decision, decisions, dog, fuck, game, games, girl, girls, grin, idea, ideas, idiot, jerk, joke, jokes, kid, kids, mistake, mistakes, motherfucker, move, movie, movies, mutt, name, ones, party, penalties, people, person, phone, piece, question, questions, reason, reasons, remark, rule, shit, smile, son, statement, statements, story, stuff, thing, things, tv	

Dumb luck, on the other hand, is a relatively well-established fixed phrase. *Lexico* provides an explanation for the phrase, that of ‘pure chance’. In COCA, a List query for *dumb luck* results in 308 hits, while *stupid luck* is found just 11 times. Of these 11 hits, one seems to be a duplicate and four refer to a programme, “Crazy Stupid Luck”. Finally, compare the concordance lines below:

- (7) Meanwhile, Sugar can't even be stopped by dumb luck, as he walks away from the car accident... (COCA, Blog, 2012, forums.utsandiego.com)
- (8) Turn a new leaf or just dumb luck? (COCA, TV, 2013, *Shameless*)
- (9) It's easy to forget that taking great photos is (give or take) 80% setting up a great shot, 15% knowing how to use your camera, 4% the camera itself, and the last 1% is just plain stupid luck. (COCA, Web, 2012, digital-photography-school.com)
- (10) And if the grace of God came into things it was stupid luck, blind luck, a car pulling to the shoulder in time. (COCA, Fiction, 1995, *Paris Review*)

At least in these examples, the function of *dumb* and *stupid* seem to be very similar and in each case, could be replaced by *pure chance* without changing the denotational meaning of the utterance. As such, it is not immediately clear why *dumb luck* is so much more common than *stupid luck*. This may be due to differences in history, sound profile, or some aspect of

meaning between the words. A deeper look into the topics will not be provided in this study, as they fall outside of its scope.

Among the common collocates, words that prefer either *dumb* or *stupid* were identified as detailed in section 3.3. Six collocates preferring *stupid* and eight for *dumb* were found. Those for *stupid* were *idiot*, *party*, *game*, *boy*, *cow*, and *girl*. The *dumb*-preferring collocates were *animal*, *ass*, *animals*, *fuck*, *bastards*, *kids*, *move*, and *kid*. All collocates were listed in order of MI-score difference, where *animal* had the highest difference of 2.99 and *kid* the lowest of 0.98, both in favour of *dumb*. These collocates of *stupid* are quite a lot more neutral on their own than those of *dumb*, with *idiot* being the only offensive word. With *dumb*, only *animal(s)* and *move* are purely neutral, whereas for *kid(s)* the case is not quite as straightforward, as it can be used affectively with either positive or negative nuance.

However, while the words themselves may be neutral, *stupid* and *dumb* ‘paint’ them in a negative light. Take *cow*, for example:

- (11) Murdered your housekeeper? Don't bullshit me; I work in the Industry that invented it you stupid COW! (COCA, Fiction, 2004, *Cellular*)
- (12) Why are you making such a big deal about this? It's just some stupid cow milking. (COCA, Movie, 1996, *The Secret World of Alex Mack*)
- (13) His word against fucken hers. Stupid cow told him if he ditches the Tesco's job he won't get benefits for three months. (COCA, Fiction, 2013, *Overland*)

In (11) and (13), the referent of *stupid cow* is clearly a person, while (12) exhibits a more expected behaviour of describing something as annoying or boring. In (11), the person is only known to be some person, whereas in (13) the referent is clearly female, as *stupid cow* could be replaced with *she*. This kind of construction exhibits a negative semantic prosody, as it first evaluates the collocate *cow* as stupid and then makes a reference to a person, usually a woman that the utterer probably is not fond of. This seems to be an overall pattern, although in many cases the referent only becomes clear from the larger context.

In addition to *cow*, *animal(s)* expresses various meanings when modified by *dumb*. Compare the following concordance lines:

- (14) Later, he distinguishes between mute creatures and dumb animals; among the latter, dogs... (COCA, Academic, 1997, *Raritan*)
- (15) Mary pictured the dumb animals always running and standing, running and standing. (COCA, Fiction, 1993, *Southern Review*)
- (16) Our troops are dumb animals being led to the slaughter? (COCA, Blog, 2012, neoneocon.com)
- (17) In some species, especially elephants, great apes, and marine mammals, the old phrase " dumb animal " borders on heresy. In fact, the line between human and animal intelligence is fading, fast. (COCA, Spoken, 2009, *CBS_SunMorn*)
- (18) It is extremely hard to avoid a deer that darts out into the road in front of you and that doesn't make you a shitty driver if you hit a deer, it makes you unlucky enough to have a dumb animal dart in front of your car. (COCA, Blog, 2012, skepchick.org)

Example (14) was hand-picked as an example of the ‘expected’ use of *dumb animal*. In this context, a *dumb animal* can produce sound, but not human sound, whereas a *mute creature* cannot produce any sound. Lines (15), (17) and (18) refer to the sense of ‘lacking intelligence’ of *dumb* and relate this to animals, which is made especially clear in (17) and (18). On the other hand, (16) likens *dumb animals* to army troops. It asks the question whether the troops the writer is referring to are like the unintelligent animals described above. This could refer to herd-mindedness, the blind following of orders, or some other feature, but certainly not muteness.

One factor that could support *dumb* being categorised as American is the number of unique collocates that are modified by *dumb* relatively frequently. For this, percentages of how often *dumb* or *stupid* collocate with a word out of all instances of collocation of that word were examined. In this case, the threshold for these kinds of collocates to be considered significant is 1% of all instances of that word collocating with either *dumb* or *stupid*. *Stupid* only has one of these collocates, while *dumb* has eight: *prank* for *stupid*; *cluck*, *jock*, *shits*, *fucks*, *jocks*, *hick*, *asses*, and *redneck* for *dumb*. According to *Lexico*, the words *cluck*, *jock*, *hick*, and *redneck* are all primarily American English. In addition, the *dumb*-unique *fucks* is supported by the *dumb*-preferring *fuck*. This suggests that *dumb fuck(s)* is preferred to *stupid*

fuck(s) in American English. A more interregional study could provide more insight whether this preference is only found in the American variant or if the regions agree.

Another interesting finding from the unique collocations is the amount of collocates that directly or indirectly refer to people. Where the number of collocates of *stupid* referring to people is six, that same number for *dumb* is eighteen. These words can range from neutral, like *man*, *woman*, and *brother* with *stupid*, to very offensive, such as *slut*, *nigger*, and *motherfuckers* for *dumb*. In total, the unique collocates of *dumb* seem to be directly more offensive against people than those of *stupid*. This difference may again suggest that *dumb* and *stupid* differ in the intensity of the insult. On the other hand, it might indicate that while both words are inherently quite informal, *stupid* could be seen as less so than *dumb*. This notion gets further support from the data on news language, a not-too-informal register, where *stupid* was heavily favoured over *dumb*.

The final point of consideration in this section is collocate groups. Loosely following the principle of semantic preference and the categorisations detailed in section 3.3, the groups were constructed using the top 50 collocates of each word. All groups were common to both adjectives. One of the most noteworthy groups was that of generalised items or groups. This group contained items such as *thing(s)*, *stuff*, *ones*, *animal(s)*, *piece*, *shit*, *people*, and *person*. Another quite notable common group was that of remarks: *question(s)*, *comment(s)*, and *joke(s)*. The remaining groups were abstract results of mental work (*idea(s)*, *decision(s)*, *reason*, *move*), (in)direct references to people (*bitch*, *ass(es)*, *girl*, *idiot*, *kid(s)*, *son*, *fuck(s)*, *bastard(s)*, *boy*, *person*, *people*, *motherfucker*, *cow*, *Americans*, *man*, *woman*, *guy*, *jock(s)*, *friends*, *waiter*, *cop*), insults (underlined in the list on people words), creations (*game*, *video*, *story*, *book*, *song*), and animals (*dog*, *cow*, *animal(s)*, *bunny*, *beast*).

Furthermore, the groups of general items, references to people, and animals were divided into more specific groups. The first group was divided into two subgroups: words

which are too general to categorise, and words which specify some category in which they belong. The first subgroup includes *thing(s)*, *stuff*, *ones*, *piece*, and *shit*. The rest of the items of the first main group make subgroup two. The group of people references was divided into foolish people (*ass(es)*, *idiot*); people in close relationship (*kid(s)*, *son*, *friends*); occupational people (*waiter*, *cop*); people differentiated by sex (*bitch*, *girl*, *son*, *boy*, *cow*, *man*, *woman*, *guy*); people in general (*person*, *people*); people who are deemed unworthy of respect (*bitch*, *ass(es)*, *idiot*, *fuck(s)*, *bastards(s)*, *motherfucker*, *cow*); people of a certain nationality (*Americans*); and people characterised by interest or success in something (*jock(s)*). The groups of animals was divided into pets (*dog*, *bunny*); farm animals (*cow*); wild animals (*beast*); and animals in general.

The group-internal variance is significant primarily in single collocates and less so categorically. For example, *animal(s)* was much more frequent with *dumb*, while *cow* was with *stupid*. *Cow*, of course, belongs to insults, sex-differentiators, and animals. *Dumb* modified words in the insult group slightly more often, whereas *stupid* modified the sex-differentiative words slightly more frequently. Animal words other than the above did not show significant preference with one word over the other. However, while many insults were more common with *dumb*, *idiot* and as above, *cow*, were more frequent with *stupid*. Based on these results, it is relatively safe to say that no one whole category is strongly preferred with *dumb* or *stupid*, but that these preferences are mostly seen in single collocates. This leads to the conclusion that the adjectives mainly differ in specific contexts and their own preferred constructions.

4.3 The case of *dumb* and *stupid ass*

As already mentioned in the previous section, both *dumb ass* and *stupid ass* are common collocations. *Ass* is the most frequent noun collocate of *dumb* and the eleventh most frequent

for *stupid*. Comparing the frequencies between the adjectives, *dumb ass* (481 instances of collocation) is more than two times as frequent as *stupid ass* (205 instances), even though *dumb* is a much rarer word. This also leads to one of the highest MI-score differences of 2.78. This suggests that there is some use unique to *dumb*. According to the *Merriam-Webster* entry for *ass*, it has three noun senses: ‘a hardy animal similar to a horse’; ‘a stupid or otherwise unbearable person’; and ‘buttocks, often also referring to a specific person.’ In addition to these, there is an adverbial sense of a postpositive intensifier. Example concordance lines are provided below, starting with *stupid ass*:

- (19) God dammit, Wendel. Who gives a fuck about your stupid ass boom box? (COCA, Movie, 2014, *Chubbies*)
- (20) That's why your stupid ass is sitting where you sitting. (COCA, Movie, 2004, *Doing Hard Time*)
- (21) "Ben, you're an ass, a dum' black stupid ass!" (COCA, Fiction, 1996, *TellMeTale*)
- (22) the Obamanaics believe American women will like being told that if your child got pregnant as a teenager, you are a stupid ass of a mother and can not do anything important, as a result, for America. (COCA, Blog, 2012, *proteinwisdom.com*)
- (23) That's why I'm telling your stupid ass no. (COCA, TV, 2015, *Empire*)
- (24) I was born here, stupid ass. (COCA, Movie, 2019, *Conterpunch*)

In (19), *ass* functions as an intensifier for the *stupid* modifying *boom box*. In (20) and (23), the *ass* is used to refer to a person, and the phrase it contributes to could be substituted for a construction using *you*. Compare (23) to ‘That’s why I’m telling you no’. The denotative meaning does not change, while the connotative meaning and semantic prosody of the statement turns into a negatively evaluative one. The *stupid ass* in (21), (22) and (24) matches with the second sense in *Merriam-Webster*. In all cases, *stupid ass* could be replaced by, for example, *idiot* and the meaning would hardly change, although in (21) the harmony between the two instances of *ass* would be broken. Compare the above lines and results to the following instances of *dumb ass*:

- (25) Yeah, you've got to put up with the bad music and the \$15 drinks, but this place is stacked with top-shelf, young, dumb ass. (COCA, Movie, 2006, *The Break-Up*)
- (26) okay toms dumb ass just put on his blogs that if you were signed up after april second you cant change your profile to 1.0 because after april second 2.0 became

- the default to all new myspace users!!!! (COCA, Blog, 2012, blog.myspacemaster.net)
- (27) Get off me! Dumb ass! Fuck you! I told you to wait in the fucking bathroom! (COCA, Movie, 2019, *Burn*)
- (28) You just forgot your keys, you dumb ass. (COCA, Movie, 1998, *The Players Club*)
- (29) Nothing! That's the same thing your dumb ass get out of a gang... nothing. (COCA, TV, 1990, *In Living Color*)
- (30) Go on, get your dumb ass outta here, I'll haul it down there later. (COCA, Fiction, 2015, *Confrontation*)

The *dumb ass* in (25) refers to buttocks more literally, as here it likely means ‘sexual partners.’ The lines (26), (29), and (30) function similarly to (20) and (23), referencing a person. Here, though, (30) is somewhat different in that the whole phrase, *your dumb ass*, is unneeded to convey the message of the clause, which is to get out of wherever the interlocutors are. One alternative to the phrase would be *yourself*, if the speaker wanted to retain the personal reference while evoking less evaluative meaning. The concordance lines (27) and (28) exhibit similar use of *ass* as (21), (22), and (24), and the two adjectives could be interchangeable or left out in (21), as it already contains *dum*. This doubling down on stupidity seen in (21) could be interpreted as a difference in intensity, which will be looked at in the next section.

In phrases where a person is called an *ass* of either the *stupid* or *dumb* kind, *ass* is the head word of the noun phrase. This is one possible reason for the spread of the word *dumbass*. Using the Chart tool in COCA and querying “dumbass”, the word has undoubtedly been spreading the periods 1990-1994 (nine instances) and 2015-2019 (122 instances). *Stupidass* is also found in the corpus, but only seven times. This difference in growth is likely due to the shorter and easier-to-say nature of *dumb*. Additionally, some instances of either *dumbass* or *stupidass* may be typographical errors of some other construction of *ass* and either *dumb* or *stupid*, or vice versa. Compare the below concordance lines with (21), (22), (24), (27), and (28) above:

- (31) "If you believe that, you a dumbass n*gga," she retorts (COCA, Magazine, 2018, *The Verge*)

- (32) Go ahead and kill yourself, stupid tar breath! Dumbass! Get outta here! (COCA, TV, 2002, *South Park*)
 (33) Cleveland, you dumbass. (COCA, TV, 2012, *The Cleveland Show*)
 (34) I'm a dumbass, yes. I'm a big, stupid dumbass. (COCA, Movie, 2018, *Party Hype*)

Example (32) is syntactically very similar to (27), where *dumbass* and *dumb ass* are in their own clauses, respectively. Their functions in the contexts seem to also mirror each other. Both phrases seem to call out some person for their actions or decisions, which have been deemed foolish by the speaker. On the other hand, (33) seems to correspond to the use found in (28). This use is only different to the one described above by the presence of *you* before the phrase, which makes the effect of the phrase more personal. In (34), both the inclusion of an article before the phrase and the combination of *stupid* and *dumb(ass)* are present. This could signal identifying as a part of a group or distancing the offensive phrase from the insulted. Finally, the use in (31) resembles that of (19), where *ass* is an intensifier. Here, it is difficult to say whether *dumbass* can also function as this intensifier, or if there has been a typographical mistake in the writing. A study more focused on investigating this phrase should provide more concrete evidence as to which case is more prevalent.

4.4 The significance of “I may be dumb, but I’m not stupid.”

As alluded to in previous sections, there is some evidence of *dumb* and *stupid* displaying varying levels of power in different contexts and uses. The two main culprits are the expression “I may be dumb, but I’m not stupid” and concordance line (3) in section 4.1.

Firstly, the above expression establishes that the speaker identifies as being at one level of foolishness, but not another. The levels in question are *dumb* and *stupid*, which are often defined as perfect synonyms, not differing in their intensity. Compare the provided expression with the below concordance lines, queried with “I may be dumb ,” and “I may be stupid ,” with the List tool:

- (35) I may be dumb, but not that dumb. (COCA, Fiction, 2006, *Iowa Review*)

- (36) I may be dumb, but I'm NOT retarded (COCA, Academic, 2003, *Education*)
 (37) I may be stupid, but I'm not bloody stupid (COCA, Movie, 2007, *Death Proof*)

Both queries resulted in seven instances of expression containing either of the above beginning phrases. A quick search for similar expressions with different modal verbs did not result in noteworthy results. Four of the seven concordance lines containing “I may be dumb,” continued with the expected “but I’m not stupid”. While the above lines appeared only once in the results, they confirm a pattern: the weaker level in the comparative expression is situated before the stronger one. This strongly suggests that *dumb* is, in some cases, considered a softer word than *stupid*.

The second point of examination is the case of *stupid* being repaired to *dumb*. Finegan (2007, 301) states that repairs are conversational tools, with which an interlocutor may correct oneself or another converser, edit something previously uttered, or repeat some passage. The current case is primarily of the second type. As previously affirmed, *dumb* is sometimes softer than *stupid*. In the case of “You’re so st... You’re so dumb”, the speaker has realised their situation mid-utterance, and decided to opt for the less offensive word choice. This kind of repair, where an expression is substituted for a less rude one, is not exactly mentioned by Finegan but could be classified as a subcategory of the editing type of repairs.

Finally, instances where both *stupid* and *dumb* are found yet again signal a difference in insulting power. This is most apparent in example (34) and is similar in logic to the weak-strong-construction described in the second paragraph of this section. The speaker first establishes that they are a dumbass, but they also add to this by modifying *dumbass* with *stupid* in the second sentence. As for (21), the phrase “dum’ black stupid ass” could have been worded without either *stupid* or *dum’*. In their current forms, the phrases are like one-two-punches, not contrasting *dumb* and *stupid*, as in the previous examples, but adding to each other.

Based on this difference in intensity, the kinds of collocates of the adjectives, especially the unique and frequent ones, become even more relevant. As established, the collocates of *dumb* are generally more vulgar. Contrast this to the current argument of *dumb* being a weaker insulting word. This dichotomy leads to the conclusion that *dumb* may be used with these offensive words partly to soften the message. This could be further supported by phonological, historical and extralinguistic factors, which would require different approaches of study.

5 Conclusion

The conclusion of the study will begin by reviewing and providing answers to the research questions. Additionally, possibilities of future study will be discussed and finally, some of the limitations of the study addressed.

The first research question was divided into two parts. The first part focused on the frequencies of *dumb* and *stupid* in American English compared to other regional variants. The primary finding was that *dumb* is greatly less common than *stupid* across all analysed regions and registers. In Ireland, Great Britain, Malaysia, and South Africa, *dumb* is used considerably less than in the other countries. In the regions apart from the four above, the frequency of *dumb* relative to that of *stupid* signals either similar or slightly smaller prevalence of *dumb* compared to the United States. There were also differences in how common in total both words were across regions. Here, the United States saw easily the most frequent use of both words, while in India, neither word is very frequent relative to the other regions. This suggests possible differences in culture or other reasons outside linguistic ones.

The second part of the first question focused on frequencies between genres in American English. *Dumb* continued to be heavily dwarfed by *stupid* across all genres. The difference in TV and movie language was the most pronounced. Where most genres favoured

stupid over *dumb* with a ratio of one to two or three, the section saw *stupid* used five times as frequently as *dumb*. The use of both words was pronouncedly low in the academic register, where informal expressions are the exception instead of the norm. The frequencies of *dumb* and *stupid* were unexpectedly low in the spoken section of COCA, especially compared to those of the TV/movies section. This is likely due to the contents of the COCA spoken section, consisting mainly of talk shows on radio or TV.

The second question involves the collocational behaviour of *dumb* and *stupid* in American English. Both words were found to have similar semantic preference and semantic prosody, the latter likely due to the nature of the words. While semantic preference was similar, there groups varied in their exact contents and the significance of specific items. This was most apparent through the Mutual Information score differences of collocates like *ass* and *idiot*. One feature similar in terms of both content and importance was the collocate group of one or more unspecified items, such as *dumb* or *stupid things*. The other collocate groups were insults, persons, remarks, animals, results of mental work, and creations. Most insults were simultaneously (in)direct references to people.

Moreover, there were significant differences in the unique collocates of the 100 most frequent one. Both adjectives collocated with thirty-six unique words. Additionally, both groups of unique words displayed differences in the words themselves. The collocates of *dumb* included more and stronger insults, like *hick*, *redneck*, and *nigger*, while the collocates of *stupid* included more denotatively neutral and impersonal words, such as *accident*, *dance*, and *job*. Concurrently, the actual functions of the adjectives in phrases like *dumb ass* or *stupid luck* seemed very similar to each other. The primary function was found to be the ‘painting’ of adjacent words in a negative semantic prosody. This semantic prosody usually coincided with connotational meanings of ‘foolishness’, ‘irrelevance’, or ‘disrespect’. Lastly, some of the unique collocates that collocated frequently with *dumb* relative to other words

were found to possibly signal it being mainly a feature of American English. These collocates included *hick*, *redneck*, *jock(s)*, and *cluck*.

The word *dumbass* has considerably risen in frequency, likely as a result of the common collocation of *dumb ass* being shortened or written abnormally. Between the beginning of the time period COCA includes data of, 1990, and the end of that period, 2019, *dumbass* has become much more significant. It is often used to replace *dumb ass* in instances like “Go home, (you) dumb ass”, but may also be used to intensify phrases in the same way that *ass* does. In contrast, *stupidass* has not become common, and is only found a few times in COCA, even though it exhibits the same meaning.

The third and final question was on semantic qualities of *dumb* and *stupid*. The investigation resulted in two main findings. Firstly, *dumb* was found a softer word than *stupid*. Phrases which contrast *dumb* and *stupid* place *dumb* as the weaker term. Additionally, *stupid* was found to have been repaired to *dumb*, but the flip case was not. This seems sensible in the context of repairs, as one would hardly choose to be more strongly insulting than vice versa. Based on this, the unique collocates of *dumb* were considered again and compared to the weaker intensity of the adjective. It was seen that *dumb* is used to contrast offensive collocates with a softer modifier to lower the intensity of the insult as opposed to the more common *stupid*. Whether this is a feature more prominent in American English than other Englishes was not determined.

Multiple suggestions for future studies have already been made in this study where they have been relevant. This thesis only analyses corpus data and this could be complemented by methods of cultural or phonological study. Alternatively, the mainly text-based approach used in this study could be accompanied with interactive methods involving people, such as interviews or questionnaires. In addition, the present corpus data is compiled of three corpora, two of which were only used for general frequencies of the adjectives. In the future, the

findings of this study could be added to by comparing them with analysis of historical or regional corpus data. The focus of a similar study could also be shifted from a large list of collocates to closer analysis of particular collocates or groups of collocates. And finally, the combination of insulting adjectives could be expanded other with near synonyms like *idiotic*, *silly*, or *foolish*.

Three primary limitations were identified for this thesis. Firstly, the restricting of collocation to the immediate right (+1) of the node word. This was done to ease the examination of words modified by *dumb* and *stupid*. This choice resulted in some obscurity of how the adjectives function in more complex noun phrases. Secondly, the comparative and superlative forms of the adjectives were not included in the analysis. This was deemed necessary in order to analyse a larger dataset. Additionally, while finishing the thesis, it was noted that the MI-scores on COCA seem to have changed since the acquisition of the data. While this did not affect the analysis of the thesis, as the relevant data had been saved beforehand, this will undoubtedly complicate comparing some of the results of this study with future ones.

Works Cited

Primary sources

Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA)*. Accessed January 15, 2021-April 27, 2021, <https://www.english-corpora.org/coca/>.

Davies, Mark. 2013. *Corpus of Global Web-Based English*. Accessed April 23, 2021, <https://www.english-corpora.org/glebe/>.

Davies, Mark. 2016-. *Corpus of News on the Web (NOW)*. Accessed April 23, 2021, <https://www.english-corpora.org/now/>.

Secondary sources

Butterfield, Jeremy, ed. 2016. *Fowler's Concise Dictionary of Modern English Usage*. 3rd ed. Oxford: Oxford University Press. Accessed April 22, 2021. <https://www-oxfordreference-com.libproxy.tuni.fi/view/10.1093/acref/9780199666317.001.0001/acref-9780199666317-e-1103>.

Cambridge Advanced Learner's Dictionary & Thesaurus, s.vv. "dumb, stupid," accessed April 22, 2021, <https://dictionary.cambridge.org>.

Collins COBUILD Advanced English Dictionary, s.vv. "dumb, stupid," accessed April 22, 2021, <https://www.collinsdictionary.com>.

Finegan, Edward. 2007. *Language: Its Structure and Use*. 5th ed. Boston, MA: Thomson Wadsworth.

Garner, Bryan. 2016. *Garner's Modern English Usage*. 4th ed. Oxford: Oxford University Press, Incorporated. Accessed April 22, 2021. ProQuest Ebook Central.

Gries, Stefan Th. 2009. "What is Corpus Linguistics?." *Language and Linguistics Compass* 3, no. 5 (September): 1225-1241. <https://doi-org.libproxy.tuni.fi/10.1111/j.1749-818X.2009.00149.x>

Kara, Abdurrahman. 2017. "A CORPUS-BASED APPROACH TO DICTIONARY DEFINITIONS IN THE INTERFACE OF SEMANTIC PROSODY AND NEAR-SYNONYMY." *Journal of International Social Research* 10, no. 50 (June): 97-103. <https://doi-org.libproxy.tuni.fi/10.17719/jisr.2017.1643>.

Kjellmer, Göran. 2003. "Synonymy and corpus work: On almost and nearly." *ICAME Journal* 27 (April): 19-27. <http://icame.uib.no/ij27/kjellmer.pdf>.

Kreidler, Charles W. 2013. *Introducing English Semantics*. London: Taylor & Francis Group. Accessed April 27, 2021. ProQuest Ebook Central.

Lexico.com, s.vv. "dumb, stupid," accessed April 22, 2021, <https://www.lexico.com>.

- Liu, Dilin. 2010. "Is It a Chief, Main, Major, Primary, or Principal Concern?: A Corpus-Based Behavioral Profile Study of the Near-Synonyms." *International journal of corpus linguistics* 15 (1): 56–87. <https://doi-org.libproxy.tuni.fi/10.1075/ijcl.15.1.03liu>.
- Merriam-Webster.com Dictionary*, s.vv. "dumb, stupid," accessed April 21, 2021, <https://www.merriam-webster.com>.
- Meyer, Charles F. 2002. Preface to *English Corpus Linguistics: An Introduction*, xi-xvi. Cambridge, UK: Cambridge University Press. <https://search-ebSCOhost-com.libproxy.tuni.fi/login.aspx?direct=true&AuthType=cookie,ip,uid&db=e000xww&AN=112619&site=ehost-live&scope=site>.
- Partington, Alan. 2004. "'Utterly Content in Each Other's Company': Semantic Prosody and Semantic Preference." *International Journal of Corpus Linguistics* 9 (1): 131–56. <https://doi-org.libproxy.tuni.fi/10.1075/ijcl.9.1.07par>.
- Taylor, John R. 2003. "Near synonyms as co-extensive categories: 'high' and 'tall' revisited." *Language Sciences* 24, no. 3 (May): 263-284. [https://doi-org.libproxy.tuni.fi/10.1016/S0388-0001\(02\)00018-9](https://doi-org.libproxy.tuni.fi/10.1016/S0388-0001(02)00018-9)
- Weisser, Martin. 2016. *Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis*. Hoboken: John Wiley & Sons, Incorporated. Accessed April 27, 2021. ProQuest Ebook Central.
- Xiao, Richard, and Tony McEnery. 2006. "Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective." *Applied Linguistics* 27 (1): 103–29. <https://doi-org.libproxy.tuni.fi/10.1093/applin/ami045>.

Appendix 1: *Stupid* and *dumb* collocate top 35 lists as retrieved from COCA interface.

<i>stupid</i>						<i>dumb</i>					
POS	COLLOCATE	FREQ	ALL	%	MI	POS	COLLOCATE	FREQ	ALL	%	MI
1	THINGS	1123	631346	0.18	6.24	1	ASS	481	47666	1.01	10.29
2	THING	1001	572448	0.17	6.22	2	LUCK	307	50805	0.6	9.55
3	PEOPLE	695	1784507	0.04	4.05	3	THINGS	272	631346	0.04	5.74
4	QUESTION	512	299919	0.17	6.18	4	QUESTION	256	299919	0.09	6.73
5	IDEA	479	264827	0.18	6.26	5	IDEA	225	264827	0.08	6.72
6	SHIT	357	105298	0.34	7.17	6	SHIT	209	105298	0.2	7.94
7	BITCH	343	32138	1.07	8.83	7	THING	208	572448	0.04	5.5
8	STUFF	339	152978	0.22	6.56	8	PEOPLE	163	1784507	0.01	3.5
9	QUESTIONS	302	178440	0.17	6.17	9	FUCK	121	89312	0.14	7.39
10	MISTAKE	207	50702	0.41	7.44	10	QUESTIONS	107	178440	0.06	6.22
11	ASS	205	47666	0.43	7.51	11	BITCH	106	32138	0.33	8.68
12	MAN	188	744223	0.03	3.42	12	STUFF	102	152978	0.07	6.37
13	GAME	185	311392	0.06	4.66	13	KID	96	110084	0.09	6.76
14	GIRL	174	198674	0.09	5.22	14	GUY	85	264746	0.03	5.32
15	DOG	158	98915	0.16	6.09	15	ANIMAL	76	58407	0.13	7.34
16	IDIOT	144	19629	0.73	8.28	16	MOVE	74	230984	0.03	5.31
17	KID	142	110084	0.13	5.78	17	SON	66	188548	0.04	5.44
18	PARTY	141	243702	0.06	4.62	18	DOG	63	98915	0.06	6.31
19	NAME	139	317389	0.04	4.22	19	JOCK	59	1911	3.09	11.9
20	MISTAKES	137	26260	0.52	7.79	20	BASTARD	57	11835	0.48	9.22
21	COMMENTS	133	104928	0.13	5.75	21	KIDS	57	244782	0.02	4.85
22	SON	129	188548	0.07	4.86	22	ANIMALS	55	66846	0.08	6.67
23	COMMENT	104	83666	0.12	5.72	23	FRIENDS	46	232402	0.02	4.62
24	MOVE	102	230984	0.04	4.23	24	ASSES	45	3619	1.24	10.59
25	DECISIONS	99	66142	0.15	5.99	25	MISTAKE	45	50702	0.09	6.78
26	JOKE	98	36085	0.27	6.85	26	NAME	41	317389	0.01	4
27	FUCK	97	89312	0.11	5.53	27	MOTHERFUCKER	39	8046	0.48	9.23
28	BASTARD	95	11835	0.8	8.41	28	MISTAKES	39	26260	0.15	7.53
29	JOKES	94	16402	0.57	7.93	29	IDEAS	38	93509	0.04	5.66
30	MOVIE	86	133220	0.06	4.78	30	WAITER	37	6595	0.56	9.44
31	BOY	86	175292	0.05	4.38	31	JOKES	36	16402	0.22	8.09
32	FACE	86	303428	0.03	3.59	32	JOKE	36	36085	0.1	6.95
33	IDEAS	84	93509	0.09	5.26	33	FUCKS	34	1358	2.5	11.6
34	SONG	83	85777	0.1	5.36	34	ONES	34	117097	0.03	5.17
35	PERSON	78	297540	0.03	3.48	35	SHOW	34	407346	0.01	3.37