

A Corpus Investigation: The Similarities and Differences of *cute*, *pretty* and *beautiful*

TAN HAI LY

Open Learning and Educational Support
University of Guelph
Canada
tanly@uoguelph.ca

CHAE KWAN JUNG

Division of Curriculum
Korea Institute for Curriculum and Evaluation (KICE)
South Korea

ABSTRACT

Since the introduction of computers in corpus linguistics, analyses of language has transformed into a more reliable guide to language than native speaker intuition. When disagreements between speakers' intuition arise over the meaning and usage of words, an analysis of corpora can provide further insight on the characteristics of the words in question. In this study, a corpus analysis is conducted to investigate the similarities and differences in the use of *cute*, *pretty* and *beautiful* using the Bank of English (BoE). The investigation specifically looked at the frequency, collocation, semantic preference, semantic prosody and phraseology of the adjectives. The results show that similarities found between *pretty* and *beautiful*, according to these aspects, indicate that these two words may be the most synonymous pair of the three. However, the findings suggest that *pretty* and *beautiful* are far from being completely synonymous and do not have the same usage in all contexts. The analysis demonstrates that uncertainties a speaker may have regarding language use may be clarified by referring to corpora.

Keywords: corpus; frequency; collocation; semantic preference; semantic prosody

INTRODUCTION

The basis of this study comes from discussions of physical appearance in English classes at a Korean girls' middle school in Seoul, South Korea. These discussions revolve around the meaning of *cute*, *pretty* and *beautiful*, terms that are often heard in the classroom due to the students' passion for Korean pop stars. The majority of the students believe that *pretty* and *beautiful* are the most synonymous pair out of the three adjectives, which runs counter to their EFL teacher's intuition that *cute* and *pretty* are the most similar.

To obtain an informal confirmation for the EFL teacher's intuition of the meaning of the three adjectives, native speakers of English were consulted for their opinion. However, more than half of their responses mirrored the students' claim: *pretty* and *beautiful* are more synonymous than *pretty* and *cute*. The responses of the students, the EFL teacher and native speakers are based on their own experiences with the language and may not be representative of English language use as a whole. Therefore a corpus study of the adjectives is required to obtain a more reasonable answer on this issue.

The aim of this study is to conduct a corpus study to investigate the similarities and differences in the use of *cute*, *pretty* and *beautiful* using the Bank of English (BoE). It is expected that the examination of these three adjectives according to key aspects of language will yield results that can help determine the most synonymous pair. The first section of this paper will briefly review the significance of corpus linguistics and introduce the key aspects of corpus analysis. In the second section, details of the methodology of this study will be

discussed before moving on to the third section where a corpus study will be attempted on the three adjectives followed by a conclusion about the issue based on the results.

LITERATURE REVIEW

CORPUS LINGUISTICS VERSUS INTUITION

Since the first computer-generated concordances appeared in the late 1950s, the study of language using corpora has undergone a major transformation (O’Keeffe and McCarthy 2010, Crafter and Jung 2014, Beng and Keong 2015, Khamis and Ho-Abdullah 2015). This revolution has allowed linguists to conduct empirical analyses of language by using software to observe large amounts of naturally-occurring discourse and process corpus data showing frequency, phraseology, and collocation (Hunston 2002). As a result, a corpus has become a more reliable guide to language use than native speaker intuition (Sinclair 1991, Stubbs 2001, Jung and Wharton 2012).

Hunston (2002) notes that intuition is a poor guide of language usage with respect to collocation, frequency, prosody and phraseology and explains that “although a native speaker has experience of very much more language than is contained in even the largest corpus, much of that experience remains hidden from introspection” (p. 20). Another argument against intuition comes from Sinclair (1991) who states that “[t]he problem with all kinds of introspection is that it does not give evidence about usage” (p. 39).

A corpus can provide specific examples of language usage but native speakers will create evidence that is invalid because they are not able to distinguish among various kinds of language patterning such as psychological associations and semantic groupings (Sinclair 1991). Despite these arguments against intuition, Sinclair (1991) and Hunston (2002) do not completely dismiss the value of intuition and believe that it is significant for evaluating corpus evidence. In addition, intuition may be a useful tool in language teaching according to Owen (1996) who states that intuition is an “essential and desirable aspect of language teaching which does not depend on corpus evidence for its integrity” (p. 219).

FREQUENCY

In corpus studies, an important aspect of language is the frequency of words and phrases. According to Sinclair (1991), frequency information is valuable because it indicates what the typical usage of a word is. For example, it can indicate the most frequent sense (meaning) or phraseology of an individual word or phrase (Hunston 2002).

Words that are frequent generally have more senses. Sinclair (1991) explains that “the accumulation of instances of a frequent word is not just more of the same, but ever more clear evidence of complexity” (p. 101). For words that are less frequent, Partington (1998) suggest that they have restricted use to special environments. Frequency information can also be used to find similarities and differences between synonyms. In Hunston (2002)’s comparison of synonyms *must* with *have to* and *incredibly* with *surprisingly*, she found that *have to* and *incredibly* are more frequent in spoken corpus than written corpus and that *must* and *surprisingly* are more frequent in written corpus. As a result, she suggests that *have to* and *incredibly* are less formal words than *must* and *surprisingly*.

COLLOCATION

Sinclair (1998, 2004) described four types of relations between lexical units that can contribute to the meaning of a text. Two of the relations, collocation and colligation, are

concerned with how words relate syntagmatically to each other. Collocation is the propensity of a word or phrase (node) to co-occur repeatedly with other words (collocates) within a specific span (Stubbs 1995).

Sinclair (1991) cited the work of Sinclair, Jones, and Daley (1970) to suggest that a span of collocates four words to the left and right of the node is the statistical limit of the attractive power of the node. Colligation, on the other hand, is the relation of co-occurrence between a node item and a grammatical category (e.g. verbs, quantifiers) (Sinclair 2004). For instance, in Sinclair (1998)'s study of *budge*, most of the collocates of *budge* were found to be verbs leading him to conclude that *budge* tends to co-occur (colligate) with the grammatical category verbs.

The significance of collocational information is the semantic relations it can show between the node and its collocates. By looking at a collocational list, it is possible to observe the range of associations of the node and also the semantic relations among its collocates (Hunston 2002). Any semantic sets that are identified from the collocates can consequently reveal the semantic preferences of the node.

SEMANTIC PREFERENCE

The third type of relation, semantic preference, was defined by Sinclair (2004) as “the restriction of regular co-occurrence to items which share a semantic feature” (p. 142). It is a type of semantic patterning where the node item collocates with words that are semantically-related. For example, in Partington (1998)'s study of *sheer*, it was found that *sheer* collocates with ‘volume’, ‘size’ and ‘weight’ indicating that it has a semantic preference for words that share the common semantic feature of ‘magnitude’.

The significance of identifying these semantic groups is that they can highlight the different senses of the node which in turn can contribute to its semantic profile (Hunston 2002); “provide observable evidence of the characteristic topic of the surrounding text” (Stubbs 2009, p. 125); and reveal attitudinal meanings that are associated with the node item (Hunston 2007).

However, there is a lack of a consensus to the definition of semantic preference; for instance, Partington (2004) refers to these attitudinal meanings as the semantic prosody of a unit of language. To prevent any confusion over the meaning of semantic preference and semantic prosody, this study will follow Hunston (2007) and Stubbs (2009)'s suggestion and adopt Sinclair (2004)'s description of semantic prosody.

SEMANTIC PROSODY

Semantic prosody is the fourth semantic relation Sinclair (2004) described and is concerned with the communicative function of what he called an ‘Extended Lexical Unit’ (ELU). Unlike a word, an ELU is a longer sequence of co-occurring items that includes a core (node) and an obligatory semantic prosody, which determines the function of the ELU.

An example of semantic prosody can be found in Sinclair (1998)'s study of *budge*. He explained that *budge* frequently collocates with words such as ‘refuse to’, ‘didn’t’ and ‘would not’ to form ELUs that have a discourse function of expressing frustration and irritation after failing to move something. This illustrated that the semantic prosody of *budge* determines its use over another verb (e.g. move) in such contexts. Sinclair's example suggests that semantic prosody is a feature of a sequence of words rather than just one word, and that perhaps meaning is carried by phrases instead of individual words. In this study, sequences (i.e. ELUs) involving *cute*, *pretty* and *beautiful* and their strongest collocates will be examined for similarities and differences in semantic prosody.

PATTERNS

This study will also examine the patterns and phraseologies of *cute*, *pretty* and *beautiful* through concordances in an attempt to reveal their usage. Sinclair (1991) suggests that “there is a close correlation between the different senses of a word and the structures in which it occurs” (p. 53). Hunston and Francis (2000) suggest two reasons for this association between patterns and meaning: first, there are many cases where different senses of words are “distinguished by their typical occurrence in different patterns”; and second, “words which share a given pattern tend also to share an aspect of meaning” (p. 3). Therefore it can be assumed that the more patterns two words share, the more synonymous they are in terms of usage.

METHODS

A QUICK SURVEY

The disagreement over the meaning of *cute*, *pretty* and *beautiful* between the EFL teacher and his students led to a quick survey being conducted. The results confirm that the students believe *pretty* and *beautiful* are the most synonymous pair, whereas the EFL teacher believes *cute* and *pretty* are more synonymous (Table 1).

For comparison purposes, eleven native English speakers were also surveyed to verify whether they shared the same opinion as the EFL teacher. Table 1 shows that 64 per cent of the native speakers share the same opinion as the students and selected *pretty* and *beautiful* as the most synonymous pair. The differences in the evaluation of the three terms suggest that an investigation of corpus data is required to identify the most synonymous pair of adjectives.

TABLE 1. Survey results of *cute*, *pretty* and *beautiful*

	<i>pretty</i> & <i>cute</i>	<i>pretty</i> & <i>beautiful</i>	<i>cute</i> & <i>beautiful</i>	Total
Grade 9 Korean middle school students	4 (14%)	25 (86%)	0 (0%)	29 (100%)
EFL teacher	1 (100%)	0 (0%)	0 (0%)	1 (100%)
Native speakers of English	4 (36%)	7 (64%)	0 (0%)	11 (100%)

DICTIONARY DEFINITIONS

The definitions of the adjectives *cute*, *pretty* and *beautiful* in the Collins COBUILD Advanced Learner’s English Dictionary (CCALED) and the Longman Dictionary of Contemporary English Online (LDoCEO) indicate similarities between the three words. The similarities are most salient when the first senses of *cute*, *pretty* and *beautiful* from the dictionaries are compared (Table 2).

It appears that *cute*, *pretty* and *beautiful* can be used to describe attractiveness in people and things and that attractive appears to be the superordinate for the three words. The differences between the words are subtle; for example, the definition of *cute* in both dictionaries includes the intensifier very before *pretty*, which might suggest that *cute* is a stronger expression of attractiveness than *pretty*. Another example of a subtle difference is

found in the definition of *beautiful* where intensifiers *very* and *extremely* precede *attractive*. This may indicate that *beautiful* is a stronger adjective than both *pretty* and *cute* for evaluating attractiveness.

TABLE 2. The first senses of *cute*, *pretty* and *beautiful* from the CCALED and the LDoCEO

	CCALED	LDoCEO
<i>cute</i>	Something or someone that is cute is very pretty or attractive, or is intended to appear pretty or attractive If you describe someone, especially a girl, as pretty, you mean that they look nice and are attractive in a delicate way	very pretty or attractive
<i>pretty</i>		a woman or child who is pretty has a nice, attractive face
<i>beautiful</i>	A beautiful person is very attractive to look at	Someone or something that is beautiful is extremely attractive to look at

The differences in the evaluation of the three terms suggest that an investigation of corpus data is required to identify the most synonymous pair of adjectives. The differences between the three adjectives are more apparent when looking at their other senses.

For instance, the second and third senses of *cute* are not listed for the other adjectives and thus, distinguishes *cute* from the rest. Regarding *beautiful*, its third sense indicates that it can be used to describe an action (e.g. beautiful shot) and may have a wider range of use than *cute* and *pretty*. This view is shared by Wolfson (1984) who explains that *beautiful* has a broader range due to its “metaphoric extension” (p. 239).

The dictionary definitions show that there are similarities and differences in usage among the three adjectives; nevertheless, the definitions do not clearly indicate which two words are the most similar. Therefore, an examination of the three adjectives according to the language aspects mentioned above is required to provide greater clarity on the issue.

CORPUS: THE BANK OF ENGLISH (BOE)

The data for this investigation is drawn from the 450-million-word Bank of English (BoE) corpus, created by COBUILD at the University of Birmingham. The BoE consists of twenty sub corpora (listed in Appendix A) that represent British English (71 per cent), North American English (21 per cent) and Australian English (8 per cent); 86 per cent of the corpora are written and 14 percent transcribed spoken data (Moon 2010). According to Hunston and Francis (2000), Sinclair (1991) and Moon (2010), a corpus can never be big enough. The use of a large corpus such as the BoE allows for greater accuracy of frequency information and the identification of central and typical usages of words.

Concordance lines were drawn from the BoE and analysed for collocation, semantic preference, semantic prosody and phraseology in this study. Sinclair (1999 cited in Hunston 2002, p. 52) advocates a method of observing patterns that involves selecting 30 random lines at a time and noting the patterns in each set until no new patterns are found. Hunston (2002) describes an adaptation of this method called ‘hypothesis testing’ where hypotheses about patterns formed from observing concordance lines are tested through further searches.

This method allows the corpus user to save time by eliminating the need to “examine every one of thousands of lines to obtain a reasonably accurate picture of how a word

behaves” (Hunston 2002, p. 52). In this study, sets of 100 random lines were examined for general patterns and 30 for specific patterns.

FINDINGS

FREQUENCY

In this study, the frequencies (i.e. frequency per million) of *cute*, *pretty* and *beautiful* were examined in the BoE. The findings in Table 3 show that *cute* is the least frequent (6.8) and *beautiful* is the most frequent (77.8) among the three words in the entire BoE.

TABLE 3. Frequencies of *cute*, *pretty* and *beautiful* in the BoE

	Occurrences	Frequency per million
<i>cute</i>	3,005	6.8
<i>pretty</i>	10,493	23.3
<i>beautiful</i>	35,004	77.8

When comparing the frequencies of the words, *beautiful* was found to be over eleven times as frequent as *cute* and over three times as frequent as *pretty* (Table 4). This may suggest that *beautiful* has a wider range of use, more senses and greater complexity than *cute* and *pretty*. Therefore, it is possible that *cute*, being the least frequent, might have the narrowest range of use and thus, is used in more specialised environments compared to the others.

TABLE 4. Frequency ratios between *cute*, *pretty* and *beautiful* in the BoE

	Ratio
<i>beautiful</i> : <i>cute</i>	11.4 : 1
<i>beautiful</i> : <i>pretty</i>	3.3 : 1
<i>pretty</i> : <i>cute</i>	3.4 : 1

The frequencies of the three adjectives for each sub corpus in the BoE were also compiled and listed in Appendix B. The results reveal that the three highest sub corpus frequencies for each word are in written format (Table 5). For all three adjectives, the top five frequencies include the *brmags* and *usbooks* sub corpora. However, the highest frequencies of *cute* and *beautiful* belong to the same sub corpus (*usephem*) which may suggest that these two adjectives are more similar in terms of the contexts in which they frequently appear.

TABLE 5. Top 5 highest sub corpus frequencies of *cute*, *pretty* and *beautiful*

	<i>cute</i>	<i>pretty</i>	<i>beautiful</i>
1	<i>usephem</i> (16.3)	<i>brmags</i> (49.2)	<i>usephem</i> (250.1)
2	<i>brmags</i> (14.7)	<i>brbooks</i> (28.7)	<i>brephem</i> (203.6)
3	<i>oznews</i> (11.1)	<i>sunnow</i> (28.0)	<i>brmags</i> (168.9)
4	<i>usbooks</i> (9.3)	<i>usbooks</i> (26.2)	<i>brbooks</i> (105.3)
5	<i>sunnow</i> (8.1)	<i>indy</i> (25.4)	<i>usbooks</i> (90.4)

When comparing the average frequencies of the adjectives according to sub corpus format (spoken and written), it was discovered that all three words are at least twice as frequent in the written sub corpora as in spoken (Table 6). This may indicate that *cute*, *pretty* and *beautiful* are more frequently found in formal rather than informal environments.

TABLE 6. Frequency comparison between written and spoken corpora

	Average frequency in written sub corpora (frequency per million)	Average frequency in spoken sub corpora (frequency per million)	Ratio
<i>cute</i>	7.3	3.1	2.4 : 1
<i>pretty</i>	25.3	11.9	2.1 : 1
<i>beautiful</i>	84.6	37.8	2.2 : 1

In summary, the frequency data suggest that there are similarities and differences between the three adjectives. All three words have similar written-spoken frequency ratios and the tendency to frequently appear in the same sub corpora although the similarity between *cute* and *beautiful* may be slightly greater. The largest differences were found in the frequency values of all three adjectives though the difference is smallest between *cute* and *pretty*. Based on these findings, it may not be possible to definitively conclude which two words are the most similar.

COLLOCATION AND SEMANTIC PREFERENCE

The strongest collocates of *cute*, *pretty* and *beautiful* in Appendix C, Appendix D, Appendix E were examined and classified according to grammatical class. It appears that for all three adjectives nouns account for the largest class which suggests that they often collocate with nouns (Table 7).

Pretty seems to colligate with nouns the most where 52.5 per cent of its strongest collocates are nouns while *cute* collocates with nouns the least with 27.5 per cent of its collocates represented by nouns. Overall, *pretty* and *beautiful* have the most similar distribution of collocates according to word class which may indicate that they are syntactically similar.

TABLE 7. Strongest collocates of *cute*, *pretty* and *beautiful* according to word class

	Noun	Verb	Adverb	Adjective	Conjunction	Other	Total
<i>cute</i>	27.5%	25.0%	17.5%	17.5%	5.0%	7.5%	100.0%
<i>pretty</i>	52.5%	10.0%	12.5%	10.0%	5.0%	10.0%	100.0%
<i>beautiful</i>	47.5%	15.0%	12.5%	10.0%	5.0%	10.0%	100.0%

To find the people and things (nouns) that *cute*, *pretty* and *beautiful* commonly premodify, the picture feature of the BoE's corpus software was used to compile the strongest R1 and R2 noun collocates. The results in Appendix F show that *cute* commonly premodifies 32 different nouns, *pretty* premodifies 38 nouns and *beautiful* premodifies 44 nouns. This might, again, suggest that *cute* has the narrowest range of use and that *beautiful* has the broadest range.

Appendix F also shows the noun collocates categorised into semantic sets, which can reveal the semantic preferences of the three adjectives (Table 8). *Cute* appears to have the fewest semantic preferences (five), while *pretty* and *beautiful* each have seven. Again, these results support the claim that *cute* collocates with nouns from a narrower range of semantic sets than the other two adjectives. Table 8 also indicates that there might be similarities in preferences among the three words due to some overlap in their semantic sets; for example, all three adjectives have a preference for nouns related to 'people', 'body parts' and 'generality'.

Yet, *pretty* and *beautiful* appear to be more similar with six out of seven of their semantic preferences being the same.

TABLE 8. Semantic preferences of *cute*, *pretty* and *beautiful*

	Semantic preferences
<i>cute</i>	people, body parts, generality, animals, structure
<i>pretty</i>	people, body parts, generality, art, location, nature, clothing
<i>beautiful</i>	people, body parts, generality, art, location, nature, structure

There are also differences between the preferences of the three adjectives. It appears that *cute* might be the more typical choice for describing attractiveness in ‘animals’ instead of the other two, whereas attractiveness in ‘art’, ‘location’ and ‘nature’ tends to be expressed with either *pretty* or *beautiful*.

However, when the semantic sets ‘location’ and ‘nature’ of *beautiful* and *pretty* are compared in Appendix F, it is apparent that *beautiful* has a larger semantic set for both categories which may imply that *beautiful* is more commonly used to describe those topics.

Other differences include *pretty*’s tendency to collocate with words related to ‘clothing’ (e.g. dress, clothes) and *beautiful*’s preference for ‘structure’ (e.g. house, buildings) which further distinguishes them from each other. Despite these differences between *pretty* and *beautiful*, the similarities between the two remain the strongest among the three adjectives and might suggest that they both have a similar semantic profile and hence, have greater similarity in use than with *cute*.

SEMANTIC PROSODY

The ELUs formed by *cute*, *pretty* and *beautiful* and their strongest collocates tend to have a consistent discourse function of complimenting the appearance of people and things as discussed in Wolfson (1984)’s study on speech acts. Figure 1 shows concordances from the BoE that appear to support this.

```

that I could see that he had been `a cute kid even at that geeky age"), it
voice; I just thought I had a voice--cute voice--you know, nice mellow voice.
and from the vantage point of their cute little house proceed to watch their
bag. <p> Top / Trousers <p> Geri's cute pink zip top is from trendy label
toilet humour and plenty of cutie-cute animals which should soften even the
<p> The new series of his fun-tastic, cute-tabulous show Kids Say The Funniest
in Virginia in the 17th century. <p> Cute animal characters, some lovely

must flutter a little, at so many pretty faces # attention was so much
<p> Adams: Hmm. There's a really pretty song called `Tallahassee # <p> J.
call 01723 373333. 8.WEYMOUTH: This pretty town found favour with royalty long
young and very slim. 6. FRILLS: A pretty and very wearable trend. 7. HIGH
to enjoy a rejuvenating day at the pretty Spa Hotel in Tunbridge Wells. On
on her three-week break in the pretty village of Les Issandres, near St
of small-town American chic with pretty floral chiffon pleated skirts,

through a side door into a long beautiful garden, lit by torchlight and
we can produce them. We're carrying beautiful images of Our Lord and Our Lady
the top is a gallery from whence a beautiful view is to be had of Rome." This
<p> Broachers: Boy, this is beautiful country. Bugs flying up his nose
2001 </dt> <p> THE world's most beautiful plane will take to the skies
broad-shouldered countenance, his beautiful head topped by curly hair, his
30, told The Sun: `She was my beautiful angel. I'm so proud of her. She

```

FIGURE 1. Concordances of *cute*, *pretty* and *beautiful* expressing compliment

Although a majority of the concordances show a semantic prosody of complimenting attractiveness, there are certain ELUs consisting of adjectives and their strongest noun collocates (listed in Appendix F) that serve a different function. For instance, the 189 occurrences of the sequence ‘*pretty*+boy’ found in the BoE do not appear to express ‘compliment’, but instead they seem to have the purpose of causing insult to males as illustrated in Figure 2.

of the stitches he was too much of a pretty boy before!" <hl> Bury 1 Swindon 3; Alan Bradley, Kelly is nothing but a pretty boy wuss. WHERE were the secrets in Now look what you've done." Such a pretty boy, too." Then the man's voice, stop to look at the long-haired pretty-boy who's having his photo taken, Sylvain Jacques was just a vacant pretty-boy, clad in a preposterous see-teenager like Dale, is that he is a pretty boy, and he is liable to assaults line is that Depp is a reluctant pretty boy, cursed with beauty and fuelled by critics who insist he is too much pretty boy, not enough real fighter. In aggrieved by the comparison to a pretty-boy footballer. "I've arrived!" he we are strong and you are weak, pretty boy," Turkin told him. "That's why "

FIGURE 2. Concordances of 'pretty+boy'

These concordances indicate ELUs that include 'pretty+boy' may be used to degrade the masculinity of males by comparing male subjects to females ('long-haired *pretty* boy', 'cursed with beauty') and by criticising their toughness ('pretty boy wuss', 'you are weak'). This prosody is absent in the ELUs of 'cute+boy' and 'beautiful+boy', which appear to have the typical function of 'compliment'. Additional examples of differing semantic prosodies can be found in sequences that involve 'cute+trick'.

In Figure 3, expanded concordances are shown to illustrate that the prosody of the ELU is dependent on the context in which it is used. It appears that when 'cute+trick' is employed to describe scientific achievements (examples 1 and 4) the sequence has a function of expressing praise for a 'clever' or 'brilliant' act.

Conversely, for acts that are related to politics (examples 2 and 3) and marketing (examples 5 and 6), 'cute+trick' might have a function of criticising an act that seems 'dishonest' or 'dirty'.

1. He has managed to carve down the Y chromosome from 50 million to 6 million base pairs-and is still going strong. <p> This is a cute trick," says Brown. (newsci)
 2. ...the Mississippi governor was run out by a neophyte, and the senator from Pennsylvania, Harris Wofford, who is, in fact, in the Senate, ran as the anti-incumbent--a cute trick... (npr)
 3. Swirdlow says Mr. Bush also tried to schedule one against the World Series four years ago. <p> Joe Swirdlow (Author): That's a cute trick. (npr)
 4. We are turning against considerable field resistance. With larger instruments, power adjustment will be required." It's a pretty cute trick," said Shea. (brbooks)
 5. This cute little marketing trick will mean, of course, that you, the punter, will have the inestimable privilege of paying twice for the same songs. (brmags)
 6. USUALLY, these theme albums are little more than a cute confidence trick. The idea is that we see the list of venerable names on the sleeve and are so impressed by them that the exceptional duffness of the individual tracks completely passes us by. (guard)

FIGURE 3. Expanded concordances of 'cute+0,2trick' (no or one or two words between *cute* and *trick*)

It must be noted that there are only six instances (excluding one instance that could not be interpreted) of this sequence, and as a result, it is not possible to make the claims above more robust without additional evidence. One final note, four out of the six instances of 'cute+trick' come from British corpora, which might imply that this sequence is more common in British English than it is claimed by the CCALED and the LDoCEO.

To summarise, instances of ELUs involving *cute* and *pretty* and their strongest collocates were found to have additional semantic prosodies aside from their typical compliment function. ELUs involving *beautiful* and its strongest collocates, on the other hand, did not exhibit any additional semantic prosodies. Perhaps an investigation of sequences with *beautiful* and its weaker collocates may yield different discourse functions. However, due to length restrictions such an investigation will not be conducted in this study.

PATTERNS/PHRASEOLOGY

Cute, *pretty* and *beautiful* share general patterns for adjectives; they appear before nouns as attributive adjectives and after link verbs as predicative adjectives. Another pattern shared by the three words is the tendency to be in prepositional phrases introduced by 'with' (Figure 4).

The concordances suggest that the three adjectives share similar aspects of meaning in these phrases by providing additional details related to the attractiveness of processes and objects in the main clauses.

male news chiefs want people with cute faces, cute bottoms and nothing in hit the nail on the head with the cute Sirion - one of the prettiest coolly set the wheels in motion with a cute back heel into the path of Burchill. not tugging the heartstrings with cute furry animals. Being teamed up with Hugh Grant falling in love with a cute American girl. As I've learnt at just a plain-looking teenager with a cute sparkle in her eyes. These days the relied on photo opportunities with cute pets to pep up their flagging polls, years ago, I got off with a drop-dead cute boy from Nottingham who, having who had admitted to flirting with pretty women that she was more than a <p> Fairy tales come true with this pretty costume! Outfit includes satiny is not simply a piece of paper with pretty words. It is a living document. It So get one step ahead with this pretty red satin shoulder bag by Miss in her right hand. And, with a pretty white bonnet and tailored coat, she And the hapless hero is in love with pretty English nurse Janet. My Hero has always has trouble with pretty girls. They're dismissed because in their hot pink number with pretty white flowers. <p> The bikini has a of a love affair with a beautiful spouse. According to All That be more natural. The nose, with its beautiful nostrils, rosy and tender, is packed with lively articles and beautiful photographs that revisit the folk music is presented with songs, beautiful scenery and musical artistry. The the waiter is in love with his beautiful boss lady Josepha. Hilarious fun, show you a different Harlem with beautiful mansions and town houses built go all out with animated figures in beautiful settings -- as much fun for break in regal company with this beautiful 12-oz. ceramic mug. It's

FIGURE 4. Concordances of *cute*, *pretty* and *beautiful* in prepositional phrases

It was also found that all three adjectives have the pattern ‘ADJ to-inf’ with the number of occurrences showing the same frequency trend found earlier: *beautiful* having the most instances (173) followed by *pretty* (94) and *cute* (16). Figure 5 displays a small sample of concordance lines of the pattern for each adjective.

with southern trendies who think it's cute to attend a ball here 'now we can go <p> No chance. Kidney is too cute to become a hostage to fortune, and collapsed because the lambs are too cute to eat. There have been two sets of was the experience. Ronnie was too cute to give anything away." Wynciti and computer or surf the net. While it's cute to pretend one can't, that's all it Tory whips who thought it deliciously cute to proclaim 'I've changed my vote' rock-god poses that he's too damn cute to pull off. Simon's surly guitar the results would not have been pretty to behold. Not that the exploits of her eyes with one hand # He's too pretty to keep. Throw the poor thing out that ladybirds are not just pretty to look at. 'Several species of easy to walk to. <p> The canals are pretty to sail along and the cobbles and that emotion. Apparently it's not pretty to say these things. In music you local rivals Coventry was never pretty to watch. But he added: "My that it was just 'too old and beautiful" to eat. <p> He was something It sounds as if your disease is too beautiful to give up," which forced him to visually striking movie which is beautiful to look at - even at its grisly and dying, destitutes, and it is so beautiful to see how these people, who have winter, but the flowers are too beautiful to waste. They brighten the place deep powder, good snowboarders are beautiful to watch. It is like surfing a Khan he'd have been too busy being beautiful to write books. A bit hard on

FIGURE 5. Concordances of ‘ADJ to-inf’ for *cute*, *pretty* and *beautiful*

Upon closer inspection, it was observed that a large portion of the concordances of this pattern for *pretty* and *beautiful* includes many verbs related to ‘observation’ such as ‘behold’, ‘look at’, ‘see’ and ‘watch’ which were not found with *cute*. The concordance lines with these verbs suggest that the pattern for both adjectives has a sense of ‘giving pleasure’ where the thing, person or activity being described by the pattern gives pleasure to the viewer (Figure 6). This finding supports Hunston and Francis (2000)’s claim that words that share the same pattern may also share the same sense.

them. For although butterflies are pretty to behold as they dine on nectar, sent out in such an equipage. It is pretty to observe how they regulate all camera has just been launched. It is pretty to look at, simple to use (wittily navigates. But though it was pretty to see these dots of light waving he is as the top of his game it is pretty to watch. <p> He is so smooth. <p> his subject. <p> Pierce's book is beautiful to behold. It is probably the with our production. Riverdance was beautiful to look at, but ours is not only and also thought-provoking book, beautiful to look at and entertaining to world champion Schuey said: 'It was beautiful to see the way the fans reacted. women work? Their hands are beautiful to watch. Moving quietly they

FIGURE 6. Concordances of ‘ADJ to-inf’ with ‘observation’ verbs

However, there was a difference found between *pretty* and *beautiful* related to the pattern: 39 of the 44 instances of ‘*pretty* to watch’ are preceded by a negative item (e.g. not). The addition of a negative item before *pretty* appears to change the sense of the pattern to ‘unpleasantness’ or ‘ugliness’ (Figure 7).

```

the TV set, Emily. This won't be pretty to watch. Lord, how did they get me
scenes of domestic violence aren't pretty to watch. <p> Tamahori is saying `
local rivals Coventry was never pretty to watch. But he added: `My
<p> Celtic's current approach is not pretty to watch and once the fans get used
we want to play. <p> The game wasn't pretty to watch and we dogged it out. But
for the murder of his brother. Not pretty to watch, but hard to forget. THE

```

FIGURE 7. Concordances of ‘negative+*pretty* to watch’

Other patterns of *pretty* were also discovered to share a similar sense such as ‘negative+*pretty* sight’ (170 occurrences) and ‘negative+*pretty* picture’ (46 occurrences) (Figure 8). Both ‘sight’ and ‘picture’ are strong collocates of *pretty* which may indicate that these patterns are not rare and that *pretty* is often used in phrases to describe ‘unpleasantness’ and ‘ugliness’.

```

society, but it--it was not a pretty sight--that execution. <p> Simon: I--is--
all strapped up and it's not a pretty sight," he added. <p> Over in Kerry,
ill # lung butter" isn't a pretty sight! The humour's rude but relentless...
facts here, naturists aren't a pretty sight because all, and I repeat all, are
you'll do. You won't be a pretty sight, but you'll live." Father O'Brian
by instant death. It's not a pretty picture but it's efficient. The
strapped on and it won't be a pretty picture. Tachi's dead.' Chief Inspector
hype. And what's left isn't a pretty picture. The European Community's plan

```

FIGURE 8. Concordances of ‘negative+*pretty* sight/picture’

In summary, *cute*, *pretty* and *beautiful* appear to share similar patterns, and as a result, may also share the similar senses. However, *pretty* and *beautiful* were found to share a greater number of patterns together than with *cute* which might suggest that those two are the most similar of the three.

CONCLUSION

The purpose of this study was to examine the similarities and differences in the use of *cute*, *pretty* and *beautiful* to determine the most synonymous pair of the three. The corpus investigation specifically looked at the frequency, collocation, semantic preference, semantic prosody and phraseology of the adjectives.

The results of the frequency analysis do not clearly reveal any significant similarities between any two words. Differences in frequencies are smaller between *cute* and *pretty*, but *cute* and *beautiful* have a slightly greater tendency to appear in the same contexts. For collocation, there is greater similarity in the distribution of collocates according to word class between *pretty* and *beautiful*. Also, the semantic preferences of *pretty* and *beautiful* are more similar where six of their seven preferences are the same. In terms of semantic prosody, apart from the typical discourse function of expressing ‘compliment’ that all three words share, there were no other similarities in semantic prosody found between the words. Regarding phraseology, *pretty* and *beautiful* appear to share a larger number of patterns which may imply that there is a greater similarity in usage between them. Overall, the higher number of similarities found between *pretty* and *beautiful* according to these aspects indicates that these two words may be the most synonymous pair of the three.

However, the findings suggest that *pretty* and *beautiful* are far from being completely synonymous and thus, the two words do not have the same usage in all contexts. This study

has shown that intuition, even that of an EFL teacher, may not be a reliable guide of language use and that corpora can be a valuable resource for clarifying any uncertainties a speaker may have regarding language use.

There are some limitations of this study. Preliminary searches for the adjective *pretty* have displayed mis-tagging in the BoE corpus where some concordances of *pretty* in the adverb form have been included in the results. This mis-tagging of *pretty* may affect the accuracy of its frequency figures (i.e. total number of instances and frequency per million figures for each sub corpora) and list of collocates.

Another limitation is that conclusions suggested from this study are based on interpretations of evidence of language (concordances) and thus, cannot be accepted as fact since no corpus can truly reflect the English language (Hunston 2002). Lastly, due to the length restrictions of this paper, only the strongest noun collocates of the adjectives obtained from the ‘picture’ feature of the corpus software are examined in this study.

REFERENCES

- Beng, C. O. S. & Keong, Y. C. (2015). Functional types of lexical bundles in reading texts of Malaysian University English test: A corpus study. *GEMA Online Journal of Language Studies*. 15(1), 77-90.
- Crafter, A. & Jung, C. K. (2014). A corpus investigation of the similarities and differences between *think about* and *think of*. *Korean Journal of Linguistics*. 39(2), 337-354.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*. 12(2), 249-268.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins Publishing.
- Jung, C. K. & Wharton, S. (2012). Finding textual examples of genres: Issues for corpus users. *Korean Journal of English Language and Linguistics*. 12(1), 129-148
- Khamis N. & Ho-Abdullah, I. (2015). Exploring word associations in academic engineering texts. *3L: Language Linguistics Literature, Southeast Asian Journal of English Language Studies*. 21(1), 117-131.
- O’Keeffe, A. & McCarthy, M. (2010). *The Routledge Handbook of Corpus Linguistics*. Oxford: Routledge.
- Moon, R. (2010). What can a corpus tell us about lexis? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp 197-211). Oxford: Routledge.
- Owen, C. (1996). Do concordances require to be consulted? *ELT Journal*. 50(3), 219-224.
- Partington, A. (1998). *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins Publishing.
- Partington, A. (2004). Utterly content in each other’s company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*. 9(1), 131-156.
- Sinclair, J. (1991). *Corpus, Concordances, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive Lexical Semantics* (pp 1-24). Amsterdam: John Benjamins Publishing.
- Sinclair, J. (1999). A way with common words. In H. Hasselggard & S. Oksefjell (Eds.), *Out of Corpora: Studies in Honour of Stig Johansson* (pp 157-179). Amsterdam: Rodopi.
- Sinclair, J. (2004). *Trust the Text*. London: Routledge.
- Sinclair, J., Jones, S. & Daley, R. (1970). *English Lexical Studies*. Birmingham, UK: University of Birmingham for the Office for Scientific and Technical Information.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*. 2(1), 23-55.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers.
- Stubbs, M. (2009). The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics*. 30(1), 115-137.
- Wolfson, N. (1984). *Pretty is as pretty does: A speech act view of sex roles*. *Applied Linguistics*. 5(3), 236-244.

APPENDIX A. The twenty sub corpora of the BoE

Sub corpora	Type of English	Mode
BBC radio (bbc)	British	Spoken
British books (brbooks)	British	Written
British ephemera (brephem)	British	Written
British magazines (brmags)	British	Written
British spoken (brspok)	British	Spoken
Economist (econ)	British	Written
Guardian (guard)	British	Written
Independent (indy)	British	Written
New Scientist (newsci)	British	Written
Sun/NoW (sunnw)	British	Written
Times (times)	British	Written
Business (wbe)	British	Written
Australian newspapers (oznews)	Australian	Written
NPR radio (npr)	North American	Spoken
Academic books (usacad)	North American	Written
US books (usbooks)	North American	Written
US ephemera (usephem)	North American	Written
US newspapers (usnews)	North American	Written
US spoken (usspok)	North American	Spoken
Canadian mixed corpus (strathy)	North American	Written

APPENDIX B. BoE sub corpora frequencies of *cute*, *pretty* and *beautiful*

Sub corpora	<i>cute</i>		<i>pretty</i>		<i>beautiful</i>	
	occurrences	frequency per million	occurrences	frequency per million	occurrences	frequency per million
bbc	5	0.3	68	3.7	247	13.3
brbooks	115	2.7	1,244	28.7	4,568	105.3
brephem	9	1.9	88	19.0	945	203.6
brmags	649	14.7	2,173	49.2	7,457	168.9
brspok	49	2.4	260	12.9	1,004	50.0
econ	12	0.8	150	9.5	245	15.6
guard	181	5.6	627	19.4	2,054	63.6
indy	220	7.8	713	25.4	1,997	71.1
newsci	20	2.5	73	9.2	221	28.0
sunnw	364	8.1	1,254	28.0	3,183	71.1
times	357	6.9	1,206	23.2	3,824	73.7
wbe	9	0.9	20	2.1	29	3.0
npr	137	6.2	377	17.0	1,106	49.7
usacad	9	1.4	26	4.1	134	21.1
usbooks	301	9.3	849	26.2	2,933	90.4
usephem	57	16.3	81	23.1	877	250.1
usnews	27	2.7	114	11.4	263	26.3
usspok	6	3.0	47	23.2	24	11.9
strathy	89	5.6	266	16.7	1,237	77.7
oznews	389	11.1	857	24.5	2,656	76.0
Total	3,005	6.8	10,493	23.3	35,004	77.8

APPENDIX C. Strongest collocates of *cute* according to t-score

Collocate	n	t-score
little (adj)	295	16.42242
and (c)	930	11.84389
so (adv)	185	9.914136
s (v)	435	9.780919
a (d)	805	9.615568
re (v)	126	9.506477
she (n)	180	9.248436
but (c)	260	9.071727
he (n)	305	8.704423
cute (adj)	72	8.466286
too (adv)	97	8.268876
very (adv)	103	7.84203
with (p)	282	7.494757
really (adv)	72	7.103692
they (n)	188	6.553677
body (n)	47	6.43027
cuddly (adj)	39	6.236026
guy (n)	42	6.226733
thought (v)	52	5.937074
seeks (v)	36	5.872592
kid (n)	36	5.869741
like (p)	90	5.8389
you (n)	201	5.674928
it (n)	314	5.631745
think (v)	66	5.602863
is (v)	326	5.56235
looking (adj)	41	5.432883
pretty (adj)	34	5.431633
kids (n)	34	5.409407
was (v)	261	5.379392
look (v)	47	5.363402
t (adv)	128	5.300242
looks (v)	34	5.285193
baby (n)	33	5.266472
young (adj)	43	5.255989

(n = noun, v = verb, adj = adjective, adv = adverb, p = preposition, c = conjunction, d = determiner)

APPENDIX D. Strongest collocates of *pretty* according to t-score

Collocate	n	t-score
a (d)	4364	37.93972
woman (n)	498	21.35959
girl (n)	418	19.8917
some (d)	595	19.44343
good (adj)	439	17.64846
very (adv)	455	17.49944
she (n)	620	17.06053
with (p)	1061	15.84008
face (n)	290	15.67624
and (c)	2749	14.56132
not (adv)	688	14.29573
girls (n)	218	14.15795
sight (n)	204	14.00616
boy (n)	205	13.60683
pretty (adv)	199	13.53058
but (c)	763	13.25116
young (adj)	210	12.43566
it (n)	1142	11.66725
shape (n)	129	10.94126
like (p)	303	10.46693
polly (n)	109	10.39477

picture (n)	121	10.25419
village (n)	117	10.23956
things (n)	162	9.827397
look (v)	158	9.728792
was (v)	899	9.71524
town (n)	117	9.402705
pictures (n)	87	8.777337
flowers (n)	79	8.440312
blonde (n)	73	8.386432
just (adv)	240	8.220873
horses (n)	72	8.067042
wasn (v)	88	7.883126
little (adj)	136	7.789968
women (n)	127	7.716238

(n = noun, v = verb, adj = adjective, adv = adverb, p = preposition, c = conjunction, d = determiner)

APPENDIX E. Strongest collocates of *beautiful* according to t-score

Collocate	n	t-score
most (adv)	2683	46.058
a (d)	10775	44.0797
and (c)	11010	41.77997
she (n)	2230	33.42861
is (v)	4969	33.21463
woman	1043	30.09027
very (adv)	1212	26.97988
it (n)	4124	25.36214
house (n)	876	25.173
was (v)	3461	24.3579
beautiful (adj)	546	22.43061
women (n)	740	22.29066
girl (n)	543	21.68218
young (adj)	652	21.6409
so (adv)	1374	21.3061
with (p)	2725	17.35774
s (v)	3696	16.53171
world (n)	689	16.22681
but (c)	2050	16.01785
her (d)	1133	15.7824
countryside (n)	244	15.2022
life (n)	531	14.89012
love (n)	366	14.79812
ever (adv)	373	14.76281
this (d)	1675	14.45718
things (n)	417	14.38741
gardens (n)	225	14.31414
look (v)	407	14.26701
wife (n)	293	14.21401
really (adv)	431	14.1825
place (n)	396	13.48548
city (n)	356	13.03647
are (v)	1804	13.00857
looked (v)	269	12.96078
garden (n)	223	12.87471

(n = noun, v = verb, adj = adjective, adv = adverb, p = preposition, c = conjunction, d = determiner)

APPENDIX F. Strongest R1+R2 noun collocates of *cute*, *pretty* and *beautiful*

Category	Collocates of <i>cute</i> (32)	Collocates of <i>pretty</i> (38)	Collocates of <i>beautiful</i> (44)
People	kid, kids, baby, babies, boy, guy, girl, girls, woman, characters, blonde, lesbian	woman, girl, boy, girls, women, blonde, boys, wife, teenager, female, lady, baby	woman, women, people, girl, wife, daughter, girls, baby, blonde, lady, boy, children, stranger
Body part	faces, face, bottoms, bum, butt, ass	face, faces	eyes, face, hair
Animal	animal, animals, puppy, bear	horses	
Art	comedy	picture, pictures, colours, style	music, piece, book
Generality	one, stuff	things, stuff, thing, one	thing, things, one
Location		town, village, place	place, south, country, city, places, surroundings, setting, grounds
Structure	cottage, bedroom	room	house, buildings, home
Clothing		clothes, dress, dresses, patterns	
Nature		flower, garden	garden, gardens, flowers, countryside, valley, scenery, beaches
Other	accent, doll, idea, trick, catch	penny, sight, football, stain, Polly	game, day, voice, laundrette