A Corpus-based Study of Synonymous Epistemic Adverbs *Perhaps, Probably, Maybe and Possibly*

Bei Yang
Bei Yang, School of English Education, Guangdong University of Foreign Studies, Guangzhou 510420, China
Email: bei_yang@gdufs.edu.cn

Abstract

Epistemic adverbs *perhaps, probably, maybe and possibly* are near synonyms, which share similar denotational meanings but differ in their usages. Using the 100 million-word British National Corpus (BNC) as data and the software Sketch Engine (SkE) as the analyzing tool, this study examines the usage differences among epistemic adverbs by conducting the analysis of concordance, n-grams and word sketch difference. The results show that different functions of SkE can make different contributions to the discrimination of epistemic adverbs. At the end of the paper, pedagogical implications of this study are discussed.

*Keywords:* Epistemic adverb; British national corpus; Sketch engine.

1. Introduction

Modality concerns the speaker’s “attitude” toward the content of what he is saying, including obligation, necessity, permission, volition, intention, ability, possibility, certainty, etc. Linguists usually group modality into three categories, namely dynamic modality, deontic modality, and epistemic modality (Palmer, 1986). Epistemic modality refers to possibility and necessity in the mental world as in the process of human reasoning, which can be expressed in verbs, adverbs, and other forms.

In this paper, I will examine the usage differences among epistemic adverbs *perhaps, probably, maybe and possibly* by using the 100 million-word British National Corpus (BNC) as data and the software Sketch Engine (SkE) as the analyzing tool. The rest of this paper is structured as follows. In the next section, I will give an overview of related work by introducing corpus approaches to synonyms. Section 3 introduces corpus data, corpus tool and analysis procedure used in this study. The results of this study are presented and analyzed in Section 4. The final section summarizes major findings and pedagogical implications of this study.

2. Related Work

Synonymy, or semantic equivalence, is an important yet intricate linguistic feature in the field of lexical semantics. Although synonyms share similar meanings, they differ in shades of meaning and vary in their connotations, implications, and register (DiMarco *et al.*, 1993). Any natural language consists of a considerable number of synonymous words. English is particular rich in synonyms due to historical reasons, which constitute a thorny area for EFL (English as Foreign Language) learners. As a result, an important aspect of English linguistics is to find proper measures of automatically identifying and extracting synonyms to find proper measures of substitutability of lexical items in the same linguistic context constitutes a good indicator of their semantic similarity. Gries and Otani (2010) studies the synonyms big, great and large and their antonyms little, small and tiny. Other sets of synonyms that have attracted attention include *strong and powerful*...
(Church et al., 1991), absolutely, completely and entirely (Partington, 1998), big, large and great (Biber et al., 1998), quake and quiver (Atkins and Levin, 1995), principal, primary, chief, main and major (Liu, 2010), actually, genuinely, really, and truly (Liu and Espino, 2012), and in virtue of, owing to, thanks to, as a result of, due to and because of Yang (2018).

3. Method
3.1. Corpus Data: BNC

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written (Aston and Burnard, 1998). The written part of the BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The spoken part (10%) consists of orthographic transcriptions of unscripted informal conversations and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins.

BNC is, by nature, monolingual, synchronic, general and sample-based, in that it deals with modern British English, it covers British English of the late twentieth century, it includes many different styles and varieties instead of being limited to any particular subject field, genre or register, and that it contains many samples which allows for a wider coverage of texts within the 100 million limit. The corpus is encoded according to the Guidelines of the Text Encoding Initiative (TEI) to represent both the output from CLAWS (automatic part-of-speech tagger) and a variety of other structural properties of texts (e.g. headings, paragraphs, lists etc.). Full classification, contextual and bibliographic information is also included with each text in the form of a TEI-conformant header.

3.2. Corpus Tool and Analysis Procedure

The Sketch Engine (SkE) is a leading corpus tool, widely used in lexicography, language teaching, translation and the like (Kilgarriff et al., 2004; Kilgarriff et al., 2014). It actually refers to two different things: the software, and the web service. The web service includes, as well as the core software, a large number of corpora pre-loaded and ‘ready for use’, and tools for creating, installing and managing users’ own corpora. Corpora in SkE are often annotated with additional linguistic information, the most common being part of speech information (for example, whether something is a noun or a verb), which allows large-scale grammatical analyses to be carried out.

SkE has a number of core functions: thesaurus, wordlist, concordance, collocation, n-grams, word sketches, and sketch diff. I will introduce the following relevant functions for this study: concordance, n-grams, and sketch diff.

3.2.1. Concordance

The basic method in SkE to generate concordance lines is from the simple search form, as in figure 1.

```
Figure-1. Basic concordance search form
```

Users, however, often want more control than the simple search offers. By clicking on the ‘Advanced’ button in figure 1, they see the options in figure 2. Users can then specify query type, part of speech, subcorpus, filter context or text types in the advanced options.

```
Figure-2. Advanced concordance search form of the adverb perhaps
```

If users click the button ‘Go’ in figure 2, the software will generate a number of lines in figure 3.
3.2.2. N-Grams

N-grams are also called multi-word expressions or MWEs. The N-grams tool produces frequency lists of sequences of tokens. This allows users to scan the entire corpus for N-word clusters (e.g. 1 word, 2 words, ...). This allows users to find common expressions in a corpus. For example, n-grams of size 2 for the sentence "this is a pen" are 'this is', 'is a' and 'a pen'.

Figure 4 shows the basic search form of N-grams. By clicking the ‘ADVANCED’ button in figure 4, users will see figure 5.

In figure 5, users can specify N-gram length, minimum frequency, maximum frequency, subcorpus, additional criteria, text types, etc. Having made the options and then clicked the ‘Go’ button, users will see the N-grams results in figure 6.
3.2.3. Sketch-Diff

The word sketch difference is designed for making comparisons by contrasting collocations. Figure 7 presents the simple search form of Sketch-Diff. When users want more control, they can click the ‘Advanced’ button, which will generate figure 8. There are several options available in the advance search form. If users click the ‘lemma’ option, the software compares the use of two different lemmas via their collocates. If users click the ‘word forms’ option, the software compares the use of two different word forms of the same lemma via their collocates. If users click the ‘subcorpora’ option, the software compares the use of the same lemma in two different subcorpora of the same corpus via their collocates. Moreover, users can also specify part of speech or minimum frequency in the advanced options. At last, after users click the button ‘Go’ in figure 8, the software will generate a summary-list of two synonymous words in terms of collocations as in figure 9.
4. Results and Analysis

4.1. Frequencies of Epistemic Adverbs in BNC

Concordance enables researchers to compare frequencies of synonymous words. As shown in Table 1, epistemic adverbs can be roughly divided into two groups based on frequencies: *perhaps* and *probably* belonging to the high-frequency group, and *maybe* and *possibly* belonging to the low-frequency group. Although epistemic adverbs vary in their total frequencies, they all tend to appear much more often in the written texts rather than in the spoken transcripts. In written text, the occurrences of *possibly* is the highest (87%), followed by *perhaps* (86%) and *probably* (77%), *maybe* being the lowest (69%).

Table 1. Frequencies of epistemic adverbs in BNC

|    | perhaps | probably | maybe | possibly |
|----|---------|----------|-------|----------|
| Spoken | 4,596  | 6,104    | 3,104 | 906      |
| Written | 28,984 | 20,429   | 6,914 | 6,139    |
| Total  | 33,580 | 26,533   | 10,018| 7,045    |
| Per million | 298.9 | 236.17   | 89.17 | 62.71    |

4.2. N-grams of Epistemic Adverbs in BNC

Table 2 to Table 5 list the top 50 N-grams of epistemic adverbs automatically generated by SkE. Table 2 lists the top 50 n-grams of epistemic adverb *perhaps* in BNC automatically generated by the SkE. Further examination suggests that these n-grams can be roughly divided into the following four types of word clusters:

Table 2. Top 50 N-grams of *perhaps* in BNC

| Rank | Word                   | Count | Rank | Word                   | Count |
|------|------------------------|-------|------|------------------------|-------|
| 1    | perhaps the most       | 451   | 26   | perhaps he was         | 75    |
| 2    | is perhaps the         | 283   | 27   | perhaps the most       | 70    |
| 3    | It is perhaps          | 228   | 28   | I thought perhaps      | 68    |
| 4    | it is perhaps          | 227   | 29   | perhaps we should      | 67    |
| 5    | perhaps it was         | 212   | 30   | perhaps one of the     | 67    |
| 6    | I think perhaps        | 162   | 31   | That perhaps the       | 65    |
| 7    | and perhaps the        | 161   | 32   | perhaps you could      | 64    |
| 8    | and perhaps even       | 154   | 33   | it was perhaps         | 63    |
| 9    | perhaps in the         | 148   | 34   | and perhaps more       | 62    |
| 10   | perhaps a little       | 139   | 35   | It was perhaps         | 62    |
| 11   | and perhaps a          | 136   | 36   | perhaps for the        | 61    |
| 12   | was perhaps the        | 128   | 37   | perhaps we can         | 60    |
| 13   | perhaps because of     | 128   | 38   | or perhaps the         | 59    |
| 14   | is perhaps not         | 114   | 39   | perhaps not surprising | 58    |
| 15   | is perhaps the most    | 106   | 40   | should perhaps be      | 57    |
| 16   | perhaps the best       | 99    | 41   | are perhaps the        | 57    |
| 17   | is perhaps a           | 98    | 42   | and perhaps not        | 57    |
| 18   | or perhaps a           | 94    | 43   | perhaps more than      | 55    |
| 19   | perhaps it is         | 92    | 44   | perhaps because they   | 55    |
| 20   | perhaps one of         | 87    | 45   | and perhaps also       | 55    |
| 21   | perhaps I should       | 84    | 46   | perhaps the only       | 54    |
| 22   | This is perhaps        | 82    | 47   | perhaps that was       | 54    |
| 23   | But perhaps the        | 82    | 48   | can perhaps be         | 54    |
| 24   | perhaps even more      | 78    | 49   | with perhaps a         | 53    |
| 25   | is perhaps more        | 78    | 50   | perhaps we could       | 53    |

- *perhaps* + someone (*live/you*) + modal verbs (*should/could/can*), as in examples (1) to (3)
- someone (*I*) + mental verbs (*think/thought*) + *perhaps*, as in examples (4) to (5)
- *perhaps* + superlative modifiers (*the most*)/comparative modifiers (*more*), as in examples (6) to (8)
- *perhaps* + superlative/comparative degree modifiers (*even*), as in example (9)
- *perhaps* + degree adverbs (*a little*), as in examples (10)
- *perhaps* + conjunction of causality (*because of*), as in example (11)

(1) Because we can’t generalize then I, *perhaps I should* be talking about British foreign policy or Iranian foreign policy or South African foreign policy are there generalizations we can make and say well it is similar for all governments?

(2) If you feel able to lend this item, therefore, *perhaps you could* also let me know the value that you would place on it.

(3) *So perhaps we can* bluff it out and collect software by day leaving philosophical disquiet to the troubled night.
(4) I think perhaps we need to do that as part of the planning for each year's budget.
(5) I never mentioned it because I thought perhaps Faye didn't like flowers in the house.
(6) I don't know I mean er I thought timeshare was perhaps the most difficult area.
(7) Of the four sources, statute law is perhaps the best understood and nowadays, the most extensive.
(8) What is perhaps more disturbing is that many existing IT specialists do not appear to be aware of some fundamental principles involved in designing reliable, "user-friendly" and "environment-friendly" information systems.
(9) But there is another aspect of television's curriculum that is more hidden - and perhaps even more powerful - than that contained in specific programs.
(10) It was perhaps a little early to be certain, but he thought they were probably the best things he had ever done.
(11) This fact of life is reflected in recent economic analysis of the firm, which addresses the limits of authority and the options available within firms when direct supervision of a subordinate by a superior is difficult, perhaps because of information asymmetry.

| Rank | Word                      | Count | Rank | Word                      | Count |
|------|---------------------------|-------|------|---------------------------|-------|
| 1    | is probably the           | 564   | 26   | this is probably          | 116   |
| 2    | will probably be          | 498   | 27   | probably in the           | 113   |
| 3    | would probably be         | 412   | 28   | probably due to           | 113   |
| 4    | probably the most         | 376   | 29   | are probably the          | 105   |
| 5    | would probably have       | 337   | 30   | would probably not        | 102   |
| 6    | was probably the          | 325   | 31   | probably one of the       | 102   |
| 7    | it is probably            | 309   | 32   | and probably the          | 102   |
| 8    | it was probably           | 255   | 33   | is probably not           | 94    |
| 9    | probably the best         | 212   | 34   | would probably have been  | 93    |
| 10   | It was probably           | 188   | 35   | had probably been         | 93    |
| 11   | was probably a            | 175   | 36   | he would probably         | 91    |
| 12   | probably won’t            | 170   | 37   | he was probably           | 89    |
| 13   | is probably a             | 166   | 38   | I think probably          | 88    |
| 14   | It is probably            | 165   | 39   | probably the first        | 83    |
| 15   | you will probably         | 164   | 40   | probably be a            | 83    |
| 16   | will probably have        | 163   | 41   | I would probably          | 82    |
| 17   | probably wouldn’t         | 155   | 42   | and will probably         | 81    |
| 18   | is probably the most      | 149   | 43   | probably going to         | 78    |
| 19   | it would probably         | 137   | 44   | is probably the best      | 77    |
| 20   | probably have to          | 134   | 45   | is probably more          | 77    |
| 21   | probably have been        | 132   | 46   | they are probably         | 76    |
| 22   | This is probably          | 130   | 47   | they would probably       | 75    |
| 23   | it will probably          | 129   | 48   | He was probably           | 74    |
| 24   | probably one of           | 126   | 49   | probably because of       | 72    |
| 25   | which is probably         | 120   | 50   | which was probably        | 71    |

Table 3 lists the top 50 n-grams of epistemic adverb probably in BNC automatically generated by the SkE. Further examination suggests that these n-grams can be roughly divided into the following four types of word clusters:

- probably + modal verbs (will/would/won’t/wouldn’t/have to), as in examples (12) to (16)
- someone (I) + mental verbs (think) + probably, as in examples (17)
- probably + superlative modifiers (the most/the best/the first), as in examples (18) to (20)
- perhaps + conjunction of causality (due to/because of), as in example (21) and (22)

(12) The informational specialists of the future will probably be Platonists rather than Aristotelians!
(13) There probably won’t be a vaccine available for some years, so, as yet, there’s no specific treatment.
(14) He would probably have maintained that minding his own business would never have got him anywhere, least of all starring on radio.
(15) He'd never be bank manager now, and probably wouldn't have wanted to be.
(16) If Steve doesn't come back for a few days I'll probably have to go into Palma and see the airlines and the tourist board myself.
(17) And I think probably erm there is another issue which is equally important and that is the question that it must be a location where people who develop employment wish to locate and erm and develop enterprises.
(18) The river dolphins that live in muddy water are probably the most skilled echolocators, but some open-sea dolphins have been shown in tests to be pretty good too.
(19) Gary Kelly continues to impress too, though he missed probably the best chance of the night - he broke out in the middle of the pitch, was charging down with only Rik to beat when he nudged the ball just too far at the moment he should have shot past him.
(20) The sixteenth edition was *probably the first* of the more recent editions to be widely accepted.
(21) The loss of nutrients here is *probably due to* the original clearing of the forest and subsequent uptake by the crop.
(22) This difference is *probably because of* the lower energy per pulse transmitted with the piezo electric system compared with the electrohydraulic one.

| Rank | Word                     | Count | Rank | Word                     | Count |
|------|--------------------------|-------|------|--------------------------|-------|
| 1    | maybe it was             | 98    | 26   | maybe in the             | 23    |
| 2    | I think maybe           | 75    | 27   | maybe that was           | 22    |
| 3    | and maybe a             | 48    | 28   | Or maybe you             | 21    |
| 4    | and maybe even          | 41    | 29   | or maybe even            | 20    |
| 5    | maybe we should         | 36    | 30   | maybe we could           | 20    |
| 6    | or maybe a              | 35    | 31   | maybe even a             | 20    |
| 7    | maybe you could         | 34    | 32   | you know maybe           | 19    |
| 8    | Or maybe it             | 34    | 33   | that maybe the           | 19    |
| 9    | I thought maybe         | 34    | 34   | and maybe that           | 19    |
| 10   | or maybe it             | 33    | 35   | and maybe it             | 19    |
| 11   | maybe I should          | 33    | 36   | and maybe I              | 19    |
| 12   | maybe I could           | 32    | 37   | think that maybe         | 18    |
| 13   | and maybe the           | 30    | 38   | maybe they were          | 18    |
| 14   | or maybe the            | 28    | 39   | and maybe you            | 18    |
| 15   | maybe a bit             | 28    | 40   | Or maybe he              | 18    |
| 16   | and then maybe          | 28    | 41   | maybe you can            | 17    |
| 17   | I mean maybe            | 28    | 42   | maybe a few              | 17    |
| 18   | maybe you should        | 26    | 43   | and maybe we             | 17    |
| 19   | maybe a little          | 25    | 44   | and maybe some           | 17    |
| 20   | maybe we can            | 24    | 45   | thought that maybe       | 16    |
| 21   | maybe it is             | 24    | 46   | that maybe he            | 16    |
| 22   | maybe he was            | 24    | 47   | maybe some of            | 16    |
| 23   | maybe I was             | 24    | 48   | maybe it was             | 16    |
| 24   | that maybe I            | 23    | 49   | Or maybe the             | 16    |
| 25   | maybe it would          | 23    | 50   | that maybe they          | 15    |

Table 4 lists the top 50 n-grams of epistemic adverb maybe in BNC automatically generated by the SkE. Further examination suggests that these n-grams can be roughly divided into the following four types of word clusters:

- *maybe + someone or something (we/you'll/lit) + modal verbs (should/could/can/would), as in examples (23) to (26)*
- *someone (I) + mental verbs (think/thought/mean) + maybe, as in examples (27) to (29)*
- *maybe + degree adverbs (a bit/a little/a few/some), as in examples (30) to (33)*
- *maybe + superlative/comparative degree modifiers (even), as in examples (34) and (35)*

(23) Then again *maybe we should* have visualised the scene in 1980/81 and done something about it then.
(24) 'The only problem is, my car is going in for a service, so *maybe you could* do me a favour and give me a lift there?'
(25) 'I'll see you tomorrow, at twelve o'clock, and *maybe we can* get in an hour's work before lunch.'
(26) 'Now they were unfashionable, *maybe it would* be OK to get a Filofax.'
(27) 'I *think maybe* that's right on some occasions, but the thing about this is that very much depends where we are and what situation we're in and we may actually choose to use one of the other types of behaviour.'
(28) 'Thomas seems to know him well, so *I thought maybe* he was a cousin of yours or something.'
(29) 'Allowing giving them time, *I mean maybe* this is what we should be, you know, hopefully we'd be doing this anyway, but just using what we use in our family anyway to talk to, you know, as them questions.'
(30) 'She was your type, *maybe a bit* on the young side.'
(31) 'Once a year, in September, the very fit (and *maybe a little* mad) take part in the Ben Nevis Hill Race.'
(32) 'There are no easy answers but *maybe a few* guidelines as to what might be happening, for Margaret, staring dry-eyed and forlorn into a new day, and for all the other people who have ever had that sort of feeling.'
(33) 'I imagine most of us have entered that deep valley, and *maybe some* of us are in it now.'
(34) 'If we go back to the first few minutes, or *maybe even* the first few seconds, there must have been an incredibly high density of matter near that point?'
(35) 'I can make as a good a vassal out of some faithful man as can any of my counts, and *maybe even* a better one.'
I experience of both the public and private sectors, Cuckney rode out the
can/can not/could/could not/may/might
wing pages a number of the more visible and
possibly + conjunction of causality (because of due to), as in (44) and (45)
(36) We cannot possibly accept the MacSharry proposals as they stand because they discriminate against British
and possibly also 24
is possibly the 41 43 possibly as a 23
could not possibly have 41 44 might possibly be 23
you possibly can 37 45 is possibly a 23
could possibly be 37 46 could she possibly 23
can possibly be 36 48 it could possibly 22
I can’t possibly 35 49 as you possibly 22
may possibly be 32 50 and possibly more 22

Table 5 lists the top 50 n-grams of epistemic adverb possibly in BNC automatically generated by the SkE. Further
someone or something (I/he/you/we/they/it) + modal verbs (can/can not/could/could not/may/might) + possibly, as in examples (36) to (40)
possibly + superlative/comparative modifiers (the most/more), as in (41) and (42)
possibly + superlative/comparative degree modifiers (even), as in examples (43)
possibly + conjunction of causality (because of due to), as in (44) and (45)
(36) We cannot possibly accept the MacSharry proposals as they stand because they discriminate against British
(37) As if he could possibly be interested in his assistant’s hobbies!
(38) And of course that also meant she couldn’t possibly spare the time to do her usual singing spot in the club.
(39) The factory owner may possibly be caught under the same provision, and he can certainly be caught in respect of occupier’s liability for business premises under s 1(3) (b).
(40) ’Nonsense, darling,’ Louise laughed, but she felt that it might possibly be true.
(41) The last few minutes were possibly the most tense, as both teams made chances for themselves.
(42) In the following pages a number of the more visible and possibly more successful and productive organisations are identified and described.
(43) Evidence of ritual burial goes back at least to Neanderthal man and possibly even earlier.
(44) But possibly because of his broad experience of both the public and private sectors, Cuckney rode out the storm calmly.
(45) There is some evidence that transitory dampness can occur, possibly due to this condensation, in a small number of dwellings, but there is no evidence that it has led to rot.

4.3. Word Sketch Difference of Epistemic Adverbs in BNC

The word sketch difference function of SkE allows users to visually compare and contrast synonymous words according to their salient collocational context. In written texts, the occurrences of possibly is the highest (87%), followed by perhaps (86%) and probably (77%), maybe being the lowest (69%). Since perhaps and possibly are more likely to be used in written texts, the collocational differences between these two words will be compared. Epistemic adverbs probably and maybe, by contrast, are more likely to be used in spoken transcripts, the collocational differences between these two words will be compared.

From table 6, we can see that some words only collocate with perhaps, such as worth, surprising, next, inevitable and understandable, as in (46) to (50). While some words can collocate with perhaps as well as with possibly, such as mean, find, due and so on, as in (51) to (56). Some other words only collocate with possibly, such as dangerous, meet, manage and accept, as in (57) to (60).
(46) It is perhaps worth pointing out that in the past three years we have spent some £350 million on the poorer pensioners.

(47) It is perhaps surprising that circulating concentrations of both peptides rise within 15 minutes after eating.

(48) Of the sense that perhaps everything needs rethinking, that perhaps next time I might get it right.

(49) It was a tragedy, though perhaps inevitable, that these two great peoples met in conflict.

(50) Such lack of fundamental change is perhaps understandable because, since the eighteenth century, the rate of economic growth in Britain has been slow, a rate which influenced and was influenced by the nature of the institutional and social system.

(51) I mean perhaps the point I haven't brought out, which was another enormous effect from the mixed ability teaching, or the mixed ability grouping, was the improvement in the pupils erm behaviour.

(52) What can As You Like It possibly mean to someone who had never been in love, or Hamlet to someone who has never felt 'how weary, stale, flat and unprofitable seem to me all the uses of this world'.

(53) From here she'd get to know those whom Christine had known, perhaps find out where she'd lived; enter her skin, almost as if Christine were to walk again while Lucy became the ghost.

(54) How could I possibly find mine when there were so many hundreds of them, and so many black ones?

(55) Many LDCs are apparently vulnerable to political upheavals, perhaps due to the immaturity of their political institutions.

(56) Thomas Garvine was chosen to visit the Emperor, possibly due to Erskine's influence on his behalf, and was attached to a mission about to leave for China.

(57) It was included in small amounts in tonics as a stimulant, a practice that all modern pharmacology texts agree is useless and possibly dangerous.

(58) And er I had a letter back on the Friday from Stuart to say er according to Mr 's description it was very interesting and could I possibly meet him at Station on the Saturday morning.

(59) I mean, Chrissie, are you allowed to have a wedding ceremony tha tha which says erm I'll love you for as long as I can possibly manage?

(60) If this was jewellery, she could not possibly accept it, no matter how appropriate it might be for the gown she was wearing.

From table 7, we can see that some words only collocate with probably, such as right, reflect, end, represent, cause and so on, as in (61) to (65). While some words can collocate with probably as well as with maybe, such as er, erm, next and so on, as in (66) to (71). Some other words only collocate with maybe, such as new, last, such, touch and so on, as in (72) to (75).

| Collocates | Frequency of Nouns and verbs modified by possibly | Frequency of Nouns and verbs modified by perhaps |
|------------|-----------------------------------------------|-----------------------------------------------|
| worth      | 51                                            | 0                                             |
| surprising | 45                                            | 0                                             |
| next       | 26                                            | 0                                             |
| inevitable | 20                                            | 0                                             |
| understandable | 18                                          | 0                                             |
| mean       | 18                                            | 16                                            |
| find       | 23                                            | 23                                            |
| due        | 27                                            | 30                                            |
| dangerous  | 0                                             | 11                                            |
| meet       | 0                                             | 11                                            |
| manage     | 0                                             | 10                                            |
| accept     | 0                                             | 10                                            |

| Collocates | Frequency of Nouns and verbs modified by probably | Frequency of Nouns and verbs modified by maybe |
|------------|-----------------------------------------------|---------------------------------------------|
| right      | 96                                            | 0                                           |
| reflect    | 84                                            | 0                                           |
| end        | 65                                            | 0                                           |
| represent  | 52                                            | 0                                           |
| cause      | 51                                            | 0                                           |
| er         | 24                                            | 18                                          |
| erm        | 12                                            | 11                                          |
| next       | 17                                            | 27                                          |
| new        | 0                                             | 3                                           |
| last       | 0                                             | 3                                           |
| such       | 0                                             | 3                                           |
| touch      | 0                                             | 3                                           |
(61) After all, the article was excellent and she was probably right; she usually was.
(62) This probably reflects the shorter period of monitoring rather than any protective effect of the gastroscopy on cardiac rhythm (Table 2).
(63) He'd work on that and probably end up getting a Queen's Award for Industry.
(64) These impressive alterations probably represent an important event in the development of the site, even if we cannot define precisely what preceded it, how widespread it was and why it happened.
(65) This lower rate was probably caused by incomplete documentation of pseudomelanosis coli in those with carcinoma.
(66) Would have to be er probably a lot lower, this would probably be a maximum, because, I felt I needed a bit of time, to just sort of, you know, pass information.
(67) It may not be of safety but it maybe er not the sort of thing that we want.
(68) And actually, hunting doesn't kill all that many foxes, a a hunt probably erm, kills one an, at a meet if they're lucky.
(69) Oh that's a very nice letter, I was just thinking that maybe erm Jill hasn't got her birthday card cos she's the type of person who would write and say thank you.
(70) We can exchange contracts probably next week some time.
(71) I was thinking maybe next weekend I might or meet her for the day.
(72) Perhaps we do not have to throw away the wealth of old traditions before we can enjoy the attractions of modern society and maybe new technology cannot be appreciated unless it is weighed against the security and apparent solidarity of the past.
(73) That game proved to be the turning point in the club's fortunes, and although there have been too many false dawns already in a thus far unsuccessful return, maybe last night's victory is the break Kendall has been waiting for.
(74) Maybe such herbs are natural neural tranquillizers, no doubt possessing a far more harmonious interaction with life processes than those presently produced by drug companies.
(75) I mean we've talked only about about women; we've maybe touched on class, and we've not even mentioned the position of black women and the extra discrimination that they face in our society, and in other societies.

5. Conclusion
This paper explored the usage of epistemic adverbs perhaps, probably, maybe and possibly in British National Corpus with SkE. The results show that these near synonyms differ in their n-grams patterns as well as collocational behaviors.

This study has a number of pedagogical implications. First, studies in second language acquisition have shown that native-speakers memorize not only words in isolation, but also chunks of words. These chunks are viewed as the building blocks of language, and are available to speakers as ready-made units, which therefore contribute to the fluency and naturalness of their utterances. Thus, if EFL learners want to achieve native-like fluency and accuracy, they need to learn these chunks as shown from table 2 to 7. Second, since the number of synonyms in English is huge, it would be unlikely for teachers to teach each pair of them to students. It might be more promising to teach students how to use SkE to conduct their own research.

References
Aston, G. and Burnard, L. (1998). The BNC handbook: Exploring the british national corpus with SARA. Edinburgh University Press.
Atkins, B. and Levin, B. (1995). Building on a corpus: A linguistic and lexicographical look at some near-synonyms. International Journal of Lexicography, 8(2): 85-114.
Biber, D., Conrad, S. and Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge University Press: Cambridge, UK.
Church, K. W., Gale, W., Hanks, P. and Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (ed.), Lexical acquisition: Exploiting on-line resources to build a Lexicon. Lawrence Erlbaum: Hillsdale, NJ. 115-64.
Church, K. W., Gale, W., Hanks, P., Hindle, D. and Moon, R. (1994). Lexical substitutability. In B. T. S. Atkins and A. Zampolli (eds.), computational approaches to the Lexicon. Oxford University Press: Oxford. 153-57.
DiMarco, C., Hirst, G. and Stede, M., 1993. “The semantic and stylistic differentiation of synonyms and near-synonyms.” In AAAI Spring Symposium on Building Lexicons for Machine Translation. Stanford, CA, March. pp. 114-21.
Divjak, D. and Gries, S. T. (2006). Ways of trying in Russian: Clustering behavioral profiles. Journal of Corpus Linguistics and Linguistic Theory, 2(1): 23-60.
Gilquin, G. (2003). Causative ‘get’ and ‘have’: So close, so different. Journal of English Linguistics, 31(2): 125-48.
Glynn, D. (2007). Mapping meaning. Towards a usage-based methodology in Cognitive Semantics. Unpublished doctoral dissertation. Leuven, Belgium: University of Leuven.
Gries, S. T. (2001). A corpus-linguistic analysis of -ic and -ical adjectives. ICAME Journal, 25: 65-108.
Gries, S. T. and Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. ICAME Journal, 34: 121-50.
Hanks, P. (1996). Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1(1): 75-98.

Hu, C. and Yang, B. (2015). Using Sketch Engine to investigate synonymous verbs. *International Journal of English Linguistics*, 5(4): 29-41.

Kilgarriff, A., Rychly, P., Smrz, P. and Tugwell, D. (2004). *The sketch engine*. Proc Euralex: Lorient, France.

Kilgarriff, A., Baisa, V., Buštá, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlí, P. and Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1(1): 7-36.

Liu, D. (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.

Liu, D. and Espino, M. (2012). Actually, genuinely, really, and truly: A corpus-based behavioral profile study of near-synonymous adverbs. *International Journal of Corpus Linguistics*, 17(2): 198-228.

Miller, G. A. and Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1): 1-28.

Palmer, F. R. (1986). *Mood and modality*. Cambridge University Press: Cambridge.

Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. John Benjamins: Amsterdam Netherlands.

Peirsman, Y., Geeraerts, D. and Speelman, D. (2015). The corpus-based identification of cross-lectal synonyms in pluricentric languages. *International Journal of Corpus Linguistics*, 20(1): 55-81.

Xiao, R. and McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A crosslinguistic perspective. *Applied Linguistics*, 27(1): 103-29.

Yang, B. (2018). The usage of because of words in British National Corpus. *Research Journal of Education*, 4(7): 92-100.