The lexicogrammar of *BE interested*: description and pedagogy[1]

Costas Gabrielatos

(Edge Hill University)

ABSTRACT

This paper aims to contribute to both lexicogrammatical description and language pedagogy. It examines lexicogrammatical patterns of *BE interested* in L1 and L2 speech and writing, using the BNC, ICLE and LINDSEI, as well as the relevant information in pedagogical materials (grammars and dictionaries) for intermediate and advanced learners. The methodology combines quantitative and qualitative analyses, and employs automated and manual techniques. The analysis involves multiple comparisons: pedagogical information is critically evaluated in light of L1 use, L2 use is compared to L1 use, as well as to the information in pedagogical materials, in order to establish any correlations between pedagogical input and learner use. The results reveal that, even when addressed at advanced learners, pedagogical materials provide incomplete information on the lexicogrammatical patterns of *BE interested*. L1 use shows distinct differences between speech and writing, which are not mirrored in L2 use. Overall, L2 use shows correlations with pedagogical information rather than L1 use. The results also support the inseparability of lexis and grammar.

1. INTRODUCTION

This paper aims to contribute to the growing body of "pedagogy-driven corpus-based research" (Gabrielatos, 2006, see also Römer, 2004), with a focus on pedagogical lexicogrammar; that is, research which is situated at the intersection of language description, analysis of learner language, and evaluation of pedagogical materials, with the purpose of developing a body of corpus-based lexicogrammatical information for language learners.

The motivation for a study on *BE interested* is two-fold, stemming from both its nature and treatment in pedagogical materials (grammars and dictionaries for language learners). The adjective *interested* commonly features in pedagogical grammars, partly because of its contrast with the meaning of the adjective *interesting*, and the syntactic patterns each enters in, partly because of the variety of complementation patterns of *BE interested*. However, the information in pedagogical materials is not comprehensive (even at advanced levels), and no relevant frequency information is provided in corpus-based descriptive grammars (Biber, Johansson, Leech, Conrad & Finegan, 1999; Carter & McCarthy, 2006). The study can be seen as an example of the proposed approach to pedagogy-driven corpus-based lexicogrammatical research.

This paper examines lexicogrammatical patterns of *BE interested* in L1 and L2 (learner) corpora, representing both speech and writing. The analysis will compare learner and L1 use to establish similarities and differences. The above are then contrasted with the relevant information provided in pedagogical materials to establish the extent to which the latter provide an accurate and comprehensive account of the patterns, and the extent to which it can be hypothesised that learner use is influenced by the quality and quantity of the pedagogical information.

---

2. PEDAGOGY-DRIVEN RESEARCH

This study builds on the approach in Gabrielatos (2003, 2006, 2013), which combines corpus-based research on L1 and L2 use with the critical examination of the information provided in pedagogical materials. That is, in addition to comparing the information in pedagogical materials with L1 use (e.g. Biber, Conrad & Reppen, 1994), and comparing L1 and L2 use (e.g. Granger, Gilquin & Meunier, 2013, 2015), this approach attempts to triangulate the results of the above two comparisons by seeking explanations for differences between L1 and L2 use in the information provided in pedagogical materials.

Research in second language acquisition has indicated that explicit instruction, which includes providing learners with the type of metalinguistic information found in pedagogical materials, contributes to language learning (e.g. N. Ellis, 2015: 13-14; R. Ellis, 2008: 863, 881, 883, 900-903; R. Ellis, 2015: 264-265; Ortega, 2013: 137-140). In this light, learner output can be expected to be influenced by the information in pedagogical materials, as their content greatly informs the input learners receive in class (Meunier, 2012: 113). This assumption also underlies the large body of studies critically examining the information in pedagogical materials in light of corpus evidence (e.g. Biber & Reppen, 2002; Gabrielatos, 2006; Harwood, 2005; Hunston & Francis, 1998; Kennedy, 1992; Meunier & Gouverneur, 2009; Owen, 1993; Römer, 2004). A shared finding in such studies is that pedagogical materials tend to provide partial, inaccurate, or misleading information. Another criticism of pedagogical materials concerns the mismatch between the frequency of a given structure in L1 use and its inclusion or prominence in pedagogical materials. Biber, Conrad & Reppen (1994: 171) observe that "some relatively common linguistic constructions are overlooked in pedagogic grammars, while some relatively rare constructions receive considerable attention." Although they acknowledge the importance of the difficulty and teachability of particular structures in the selection and prioritisation of content in pedagogical grammars, they also argue that the frequency of particular structures should be an equally important consideration (Biber, Conrad & Reppen, 1994: 173-174; see also Biber & Reppen, 2002: 206-207). Similarly, Leech (2011: 12-15, 24) argues that frequent items must be prioritised in language teaching, as learners can be expected to encounter, and have a need to use, these items more often. An additional reason why the presence and prominence of particular items in pedagogical materials is worth investigating is related to Hasselgren's (1994) observation that the lexical repertoire of even advanced learners tends to be restricted to "lexical teddy bears" (ibid.: 237): a relatively small group of words and expressions that learners feel comfortable to use because they have encountered and used them in the past. A similar effect can be achieved by the prominence of particular grammatical (and lexicogrammatical) items in pedagogical materials. For example, Gabrielatos (2013) found that, when compared to written L1 use, learners tended to overuse the types of conditionals specified in pedagogical materials and underuse those excluded. Overall, language items that are discussed explicitly and in some detail in pedagogical materials can be expected to be seen by learners as more salient than those only presented in examples or exercises, and, therefore, it is reasonable to expect them to be used more frequently by learners.

For the reasons outlined above, pedagogical materials are examined for the presence, comprehensiveness, and accuracy of the information they provide either directly (via explicit rules and guidelines) or indirectly (via examples and exercises). Of course, correlations between the frequency of particular patterns in L2 use and the presence/absence of relevant information in pedagogical materials should not be necessarily understood as indicating causation, as learner use is also influenced by non-pedagogical sources, such as films, TV, websites, social media, and computer games. Finally, in the approach taken here, the use of the language item in focus is examined in both native and learner corpora. Of course,

pedagogy-driven research does not preclude, and frequently requires, primary descriptive research. That is, it does not simply contrast the information provided in pedagogical materials with that provided in descriptive grammars and dictionaries, as the latter may not be corpus-based, or may not cover the item comprehensively. Simply put, pedagogy-driven corpus research can also contribute to language description and theory. Figure 1 (adapted from Gabrielatos, 2005: 5, 2013: 160) summarises the elements of pedagogy-driven corpus-based research.
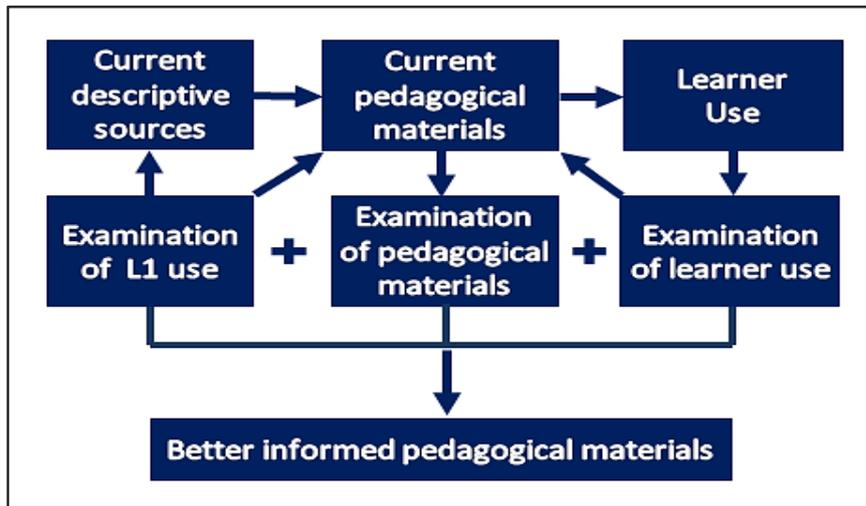

Figure 1. Pedagogy-driven research: Components and interrelations

3. CONCEPTIONS OF LEXICOGRAMMAR

Before moving to the focus and aims of this study, we must first discuss the notion of lexicogrammar, as views on its nature vary, and clarify the sense in which it is treated here. This section also provides a critical view of the primacy of lexis that is posited by most of the approaches to lexicogrammar (see Römer, 2009: 141-147). What will be argued here, and what is emerging from the analysis, is that approaches treating lexis as primary and grammar as emerging from lexical patterning (as in Hoey, 2005: 1) are not grammar-free (as acknowledged in Hunston & Francis, 2000: 1-2, 37; Stubbs, 1996: 40), nor can grammar be seen as secondary. The critical overview will start with approaches that posit the primacy of lexis, then look at Halliday's more balanced approach, and finally discuss the approach taken in this study.

Sinclair (1991: 109-115) initially discussed the "idiom principle" (exemplified by collocation) as operating alongside the "open-choice principle", or "slot-and-filler model" (words chosen according to their semantic properties to fill in particular syntactic positions), and proposed the idiom principle as accounting for "the restraints that are not captured by the open-choice model" (ibid.: 115). Sinclair (2004: 164) also distinguished between "lexicogrammar" and "lexical grammar", seeing the former as "fundamentally grammar with a certain amount of attention to lexical patterns within the grammatical frameworks; it is not in any sense an attempt to build together a grammar and lexis on an equal basis." However, his approach, later formalised as *Lexical Grammar* (Sinclair, 2004), does not treat lexis and grammar on an equal basis either, but clearly prioritises the idiom principle and a focus on lexis, with the *lexical item* as the unit of analysis. As Hunston & Francis (2000: 253) put it, lexical grammar "restored lexis in its rightful place at the centre of language description". The imbalance in favour of lexis becomes clear when we examine the components of the lexical item (Sinclair, 1996: 75; see also Stubbs, 2009: 123-126): the core (i.e. a word or

phrase), its collocates, its semantic preference, its semantic prosody, its colligations. The only component of a grammatical nature is colligation, defined as "the grammatical company a word keeps" (Hoey, 1997: 8), or "frequent co-selections of a content word and an associated grammatical frame" (Stubbs, 2002: 238). However, colligation is not an obligatory component, whereas collocation and semantic prosody are (Sinclair, 2004: 174; Stubbs, 2009: 124). Interestingly, despite the apparent prioritisation of lexis, Lexical Grammar does take account of grammar, albeit in an unacknowledged fashion. Collocation, a core component of the lexical item, is defined as relating to the co-occurrence of word-forms, as different forms of the same word can have different sets of collocates (e.g. Sinclair, 1991: 53-56). However, this can be re-stated as 'morphological marking affects collocation patterns', that is, collocation is not purely lexical, but is influenced by grammar. The primacy of lexis also underlies Hoey's work on lexical priming, which is very much in line with the tenets of Lexical Grammar. Hoey's (2005: 1) statement that "grammar is an outcome of lexical structure" clearly echoes Sinclair's (1991: 100) earlier view that "grammatical generalizations […] are the accumulation of the patterns of hundreds of individual words and phrases". However, lexical priming analysis does make use of syntactic categories (e.g. Subject, Object, Complement) (ibid.: 60).

An approach closely related to Lexical Grammar is Pattern Grammar; however, it differs from it in two important respects. First, its main focus is particular words or word classes and "the grammatical patterns they form part of" (Hunston & Francis, 2000: 1), while a pattern is seen as indicating "the behaviour of words that is typical of their word class" (Hunston & Francis, 2000: 202), and the description of particular patterns includes grammatical classes (e.g. *V of N*), and their discussion involves 'traditional' syntactic categories (e.g. passive). In this light, its main focus can be described as colligation, which is only an optional component in Lexical Grammar. Second, in this approach, grammar can also be a starting point in the analysis: "a pattern can be seen to be associated with a variety of different words" (Hunston & Francis, 2000: 43). That is, Pattern Grammar reserves a larger role for grammar than Lexical Grammar. However, it retains lexis as its main focus: grammar patterns are seen as "belonging" to a particular word (Hunston & Francis, 2000: 1-2).

Halliday proposed treating lexis and grammar as "complementary perspectives" (1991: 32), the combination of which comprises a single level of organisation in language, which he termed *lexicogrammar* (1992: 63). More precisely, Halliday conceived grammar and lexis as the notional ends of a lexicogrammatical continuum. He presented collocations and sense relations as examples of the lexis end, with polarity, mood and transitivity presented as examples of the grammar end, whereas prepositions and systems of modality were seen to occupy a middle ground (1992: 63-64). Halliday argued that the reason lexis and grammar tend to be treated separately is because "they lend themselves to different techniques of analysis" (1992: 63), in that "if you interrogate the system grammatically you will get grammar-like answers and if you interrogate it lexically you get lexis-like answers" (1992: 64). Halliday described Sinclair's approach as "tunnelling through the system interrogating it lexically while moving further and further towards the grammatical end" (1992: 64) in order to identify aspects of language use that cannot be derived from a purely grammatical analysis (1966: 410). In contrast, Halliday's approach favours grammar-like answers (1992: 64), as he perceives lexis as "delicate grammar" (1961: 267), and his aim is "to build the dictionary out of the grammar" (1992: 63). A more recent approach that has grammar as its starting point is collostructional analysis (Gries & Stefanowitsch, 2004; Stefanowitsch & Gries, 2003), which examines the lexemes that are attracted or repelled by particular slots in a given construction.

Despite refuting the traditional distinction between lexis and grammar, all the above approaches isolate particular lexical or grammatical aspects in the analysis of research

findings. That is, while discussing aspects of the lexis-grammar interaction they do focus on one or the other at different stages of the analysis. This practice is compatible with, and seems to support, Halliday's conception of the lexicogrammatical continuum described above. However, the relation between lexis and grammar is not only a matter of research angle (e.g. starting from the lexis or grammar end of the continuum), as lexis and grammar interact. Particular grammatical structures may tend to contain particular lexical items more frequently than others, or even be "lexically restricted" (Francis, 1993: 142; see also Stubbs, 1996: 40); in turn, particular lexis may be found more frequently within particular grammatical structures (e.g. Hunston & Francis, 2000: 1, Stefanowitch & Gries, 2003). Combined, the two types of interaction can be seen as instances of "lexis-grammar co-selection" (Römer, 2009: 141).

The approach taken here is that grammar is always involved (Culicover, Jackendoff & Audring, 2017), and any apparent primacy of lexis is an effect of the research focus. For example, even the seemingly lexical starting point of the present study (i.e. the word *interested*) cannot be adequately defined without recourse to grammar. That is, a full description of the starting point needs to specify that the focus is *interested* when used as an adjective, rather than as the past tense of the verb INTEREST. Similarly, BE in *BE interested* needs to be specified as a copular verb in all its tense-aspect permutations. This study approaches *BE interested* lexicogrammatically to get lexicogrammatical answers – that is, the analysis will privilege neither lexis nor grammar – or, in Sinclair's terms, neither the idiom principle nor the open-choice principle. Wherever along the lexicogrammatical continuum the research starting point may be, the exploration can move towards either side. Support for this approach is provided in a study by Erman & Warren (2000), which examined the proportion of prefabs in texts from LOB and the London Lund corpus of Spoken English (using manual analysis). Prefabs are seen as manifestations of the idiom principle, and are defined as "a combination of at least two words favoured by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalisation" (ibid.: 31). Erman & Warren (2000: 50-53) report that choices consistent with the idiom or open-choice principle have fairly similar proportions in the texts they analysed, with choices consistent with the idiom principle being slightly more frequent on average (55%) and even more frequent in spoken texts, although the proportion in individual texts ranged from 40% to 60%. They concluded that their results support their hypotheses that "in producing utterances the language user alternates between the open choice principle and the idiom principle" (ibid.: 30, 51). In this light, *BE interested* will be approached from both ends of the continuum, and the analysis will seek to derive not only grammar-like and lexis-like answers, but also answers reflecting instances of lexis-grammar co-selection. Similar examples of corpus-based lexicogrammatical research with a descriptive focus are McEnery & Xiao's (2005) examination of complementation patterns of the verb HELP, Hasselgard's (2016) examination of the patterns of '*the* N1 *of the* N2', and Rühlemann's (2007) study on the GET-passive, whereas examples of studies looking at both L1 and L2/learner use are Deshor's (2015) comparison of *to*-infinitival and gerundial complementation patterns, Gilquin's (2012) examination of the particular verbs and classes of verbs that occupy the non-finite verb slot in causative constructions, and Römer's (2009) study on the patterns of introductory *it*.

4. FOCUS AND AIMS

The main focus of this study is the complementation patterns of *BE interested,* and the extent to which relevant information about these patterns is provided in pedagogical materials. The study also examines the frequency of other copular verbs complemented by the adjective

interested, in order to establish the extent to which the complementation pattern *BE interested* is frequent enough to merit examination. An initial analysis also suggested that the focus of the study could usefully be expanded to include the following: the presence or absence of modal marking of *BE* in *BE interested*, and the types of verbs found in two complementation patterns: '*BE interested + -ing Clause*' and '*BE interested + to*-inf Clause'. At this point, it must be clarified that the study does not aim to contribute to materials design, that is, the particular presentation of information in pedagogical materials (e.g. the wording of the information, and the selection of examples), or the design of exercises. However, it is hoped that the findings can inform this aspect of pedagogical materials.

As pedagogical grammars are expected to be informed by reference grammars, three major grammars were consulted for information pertaining to the focus and aims of this study: Carter & McCarthy (2006: 21, 297, 300, 310, 439-445, 459, 462, 500, 508-509, 522, 546, 565-566), Huddleston & Pullum (2002: 50, 83, 540-544, 658, 1230-1231), and Quirk et al. (1985: 403, 413, 546, 576, 657-659, 740, 755, 774, 862, 979, 1061-1063, 1211, 1220-1292). All reference grammars mention *interested in* as a pattern – either by particular mention, or in lists of Adjective+Preposition patterns, and they all discuss the following complementation patterns of *BE interested*:

1. *BE interested in* + NP
2. *BE interested in* + *-ing* participle Clause
3. *BE interested in* + Noun (*wh-*) Clause
4. *BE interested* + *to*-infinitive Clause
5. *BE interested* + ∅ (no complementation)

The study examines the presence, absence, and prominence of information regarding the above patterns in pedagogical materials. It also examines spoken and written L1 and L2 corpora for the frequency of a) different copular verbs complementing the adjective *interested,* b) different complementation patterns of *BE interested*, and c) modalised BE in *BE interested*. Comparison between L1 use and pedagogical materials is expected to establish the extent to which the latter present L1 use and whether L1 use supports the inclusion or foregrounding of some patterns and the exclusion or backgrounding of others. Comparisons between L1 and L2 will establish the extent to which L2 use matches L1 use. Finally, comparisons between L2 and pedagogical materials will provide indications of the correlation between the information in pedagogical materials and learner use, which will be used as a proxy for determining the influence of pedagogical materials on learner use. In light of the above, the study addresses the following research questions:

In relation to pedagogical materials:

1. Do they treat *BE interested* explicitly as a pattern (i.e. do the materials reserve an entry for it)? If not, do they provide directly relevant lexicogrammatical information, that is, information on the complementation patterns of 'Copula+Adjective'?
2. Do they provide information regarding other copular verbs complemented by the adjective *interested*?
3. Do they present all complementation patterns of *BE interested* (or just *interested*, irrespective of the copular verb)? If so, are these patterns treated explicitly (i.e. is information provided directly) or implicitly (i.e. do the patterns feature in examples or exercises)?
4. What other relevant information is provided (e.g. frequencies, collostructional choices)?

In relation to use, as evidenced in the corpora examined:

5. What is the frequency of different copular verbs of which *interested* is a complement?
6. What is the frequency of each complementation pattern of BE *interested*?
7. What is the extent of modal marking of BE in each complementation pattern?
8. Are particular types of verbs used more frequently in the non-finite slots in two of the complementation patterns (BE *interested in + -ing* Clause and BE *interested + to-*infinitive Clause).
9. Are there any differences between L1 and L2 use and between spoken and written production?

5. DATA AND METHODOLOGY

The study examines two types of data: pedagogical materials (grammars and dictionaries) and corpora (L1 and L2). The L2 corpora include learner writing and speaking at CEFR levels B2-C2 (see section 5.2 for details).

5.1 Pedagogical materials

The pedagogical materials examined are at the level that the learners whose use is represented in the corpora can be expected to have used (B1-C2). It must also be clarified that the grammars used in the study are not textbooks/coursebooks (i.e. they are not meant for use in class) but reference materials for learners (like dictionaries). Learners are not expected to plough through every page and learn everything contained in the book; rather, these types of pedagogical materials are to be used when a learner needs help in understanding/using a particular language item (e.g. when writing an essay). For this reason, pedagogical materials can be comprehensive, without being intimidating for the learner. Although it is not assumed that learners will have necessarily consulted the specific pedagogical materials examined in the study, they can be expected to be largely representative of the kind of input L2 learners receive, as the content of pedagogical materials tends to be fairly uniform (e.g. Gabrielatos, 2003, 2006; Tomlinson, 2013: 2-3, 17). Table 1 lists the pedagogical materials examined (for more details, see appendix). Parentheses after the grammar titles indicate the range of CEFR levels each grammar is designed to address; links after dictionary titles lead to the entries examined.

Table 1. Pedagogical materials examined in this study

Grammars	Dictionaries
• *English Grammar in Use* (B1-B2) • *Active Grammar Level 2* (B1-B2) • *Collins COBUILD English Grammar* (B1-C2) • *Grammar and Beyond 4* (B2-C2) • *Advanced Grammar in Use* (C1-C2) • *Active Grammar Level 3* (C1-C2)	• *Cambridge Dictionary Online*: dictionary-beta.cambridge.org/dictionary/english/interested; dictionary.cambridge.org/dictionary/english/interested • *Collins Dictionary Online*: collinsdictionary.com/dictionary/english/interested • *Longman Dictionary of Contemporary English Online*: ldoceonline.com/dictionary/interested • *Macmillan Dictionary Online*: macmillandictionary.com/dictionary/british/interested • *Oxford Advanced Learner's Dictionary Online*: oxfordlearnersdictionaries.com/definition/english/interested

In the case of pedagogical grammars, the information required to address questions 1-4 was sought using the tables of contents and indexes, and the examination of relevant entries ('rules'), examples and exercises. In the case of dictionaries, information was collected by looking up the entry for *interested*: definitions, examples, and other relevant information.

5.2 Corpora

L1 use is examined via the BNC (written and spoken sub-corpora, henceforth BNCw and BNCs, respectively). Learner language is examined via two corpora: ICLE (Granger, Dagneaux, Meunier & Paquot, 2009), a corpus of essays written by upper intermediate to advanced learners of English (CEFR levels B2-C2)[2] and LINDSEI (Gilquin, De Cock & Granger, 2010), a corpus of elicited spoken English, produced by advanced learners (C1-C2)[3]. It must be mentioned that the comparison between the learner corpora and the BNC is of course not ideal, as the BNC contains a variety of genres, whereas ICLE and LINDSEI are genre-specific. Therefore, comparisons between L2 and L1 use would in principle be more usefully carried out between corpora containing similar genres (Biber et al., 1994: 174, 183; Leech, 2011: 13); that is, between the above learner corpora, on the one hand, and LOCNESS and LOCNEC, on the other. In this light, the results of the comparisons between L1 and L2 use must be seen as tentative, as the study needs to be supplemented with comparisons between genre-compatible L1 and L2 corpora. However, LOCNESS and LOCNEC would not be useful in the evaluation of the pedagogical materials, which provide non-genre-specific information. In this light, the use of a general corpus (BNC) is essential.

The BNC was accessed via BNCweb (Hoffmann, Evert, Smith, Lee, & Berglund-Prytz, 2008), ICLE and LINDSEI via two versions of CQPweb (Hardie, 2012; Xu & Wu, 2014).[4] For aims 6-8, random concordance samples of 250 instances of the word-form *interested* were analysed. Samples were retrieved using this simple query, rather than using a complex query to return instances of *interested* used with the copular BE, for two reasons. First, it enabled the manual analysis to determine the proportion of BE in relation to other copular verbs complemented by the adjective *interested*. Second, it is expected to have resulted in very high recall (excluding instances of misspelling or mistranscription of *interested* in the corpora). This was particularly useful in the case of spoken corpora, in which features of spontaneous speech (e.g. fillers, repetitions, false starts) would make it likely that true positives would be missed and false positives would be included. For example, it would be difficult to formulate a complex query that could accommodate instances like (1) below, where a hesitation (*mm*) and a pause intervene between *interested in* and the *–ing* Clause complementing it. Samples were first examined in order to remove predicative uses of *interested* (2), learner mistakes (3), elliptical uses (4), and incomplete utterances (5).

(1)   I I really I I don't know I don't have much preferences because (erm) *I'm very interested in* (mm) . *talking* about this in (eh) . (er) in literature written in English [LINDSEI IT029]

(2)   Is it not better off as a hidden surprise to be discovered by the *interested* tourist? [BNCw K5M1867]

(3)   I h= em I had in school and eh it eh is eh *interested* because . now I can do things that I want . and erm are eh eh of ver= . of interested to me . [LINDSEI GR040]

(4)   The people, who watch public television *are* in average more educated, often *interested in* politics and keenly following all daily events, they live with the rest of their society through television. [ICLE CZPU1004]

(5)   6494   Well, I mean we just said well that's, you know, an expensive place for a holiday by the sound of it, but erm we hadn't got into any details.
        6495   Well this was erm Bulgaria
        6496   Yes, yes
        6497   Mm
        6498   but er so I, I don't know if he'd *be interested in*.
        6499   I certainly wouldn't coach it <pause> so <pause> I think you would be creased by the time you got there in two days.

                                                                [BNCs KDM6498]

Frequency comparisons took into account both the size of frequency differences and their statistical significance (established using the log-likelihood test).[5] Sizeable differences were deemed those in which the frequency of a feature in one corpus was at least 25% higher than in the other; the threshold for statistical significance was set at $p \leq 0.05$ ($G^2 \geq 3.84$). It must be clarified that frequency differences will be presented in relation to the highest frequency, so that differences can be comparable. Although all sizeable differences will be discussed, only those that are also statistically significant will be deemed reliable. To facilitate reading the tables reporting on comparisons, in cases of sizeable differences, the cell of the highest frequency will be shaded; if the sizeable difference is also statistically significant, the cell indicating the G^2 value will also be shaded.

In order to address question 8, an automated collocation analysis on the whole corpora was employed, so that the strength of collocation could be established (something that could not be achieved via the manual analysis). As the analysis of copular verbs complemented by the adjective *interested* (question 5) showed that BE is used in 92%-99% of instances, collocates were calculated on the queries 'interested in' and 'interested to'. Right-hand collocates in position R1 for written corpora, and R1-R3 for spoken corpora (to allow for fillers and dysfluencies). Granted, the R1 position may exclude discontinuous constructions (e.g. As a result, economists are much more *interested than before in* correcting inefficiencies by reducing the costs of doing business [ABE 3219]; I *was interested last year to see* Tim Jonke's article and his method of spraying acrylic paint over oil paint [C89 1296]). However, such instances are extremely rare: in BNCw, queries with one and two intervening words between *interested* and *in* (followed by a verb) returned 34 and 3 instances, respectively, whereas there were 5016 instances of *interested in* followed by a verb. The corresponding results for *interested* and *to* were 7 and 9 instances, respectively, compared to 366 for *interested* followed by a *to*-infinitive. Therefore, the R1 span provides a dependable picture of collocates. The minimum frequency for both collocate and collocation was set at 1, so that even very infrequent collocates would be taken into account. For a word to be deemed a collocate, the MI score of the collocation had to be at least 3.

---

[5] Calculations of frequency differences and statistical significance were carried out using Paul Rayson's spreadsheet (http://ucrel.lancs.ac.uk/people/paul/SigEff.xlsx)

6. RESULTS: ANALYSIS AND DISCUSSION

6.1 Pedagogical Materials

Most of the pedagogical grammars examined present *interested* as an adjective, or make students aware that words ending in *–ed* can be adjectives (five out of six), and mention that *-ing* forms can function as nouns (four out of six). Dictionaries, due to their focus, do not provide any general grammatical information regarding *-ed* and *-ing* forms; however, they all contain an entry or sub-entry on *interested* as an adjective. When discussing complements of *interested in*, pedagogical materials (particularly dictionaries) rarely distinguish between *-ing* forms syntactically functioning as nouns and verbs, but tend to focus on reminding learners that a word with the suffix *-ing* must follow *interested in*. More to the point, they do not make it explicit that an *-ing* form may belong to a different part of speech according to the complementation pattern (and corresponding syntactic slot): a noun in 'BE *interested* in + NP' or a verb in 'BE *interested in* + *-ing* Clause'). Not all pedagogical grammars present *interested in* as a pattern, whereas most dictionaries (four out of five) explicitly include it in their entry for *interested*, and all provide examples with *interested in*. Only two grammars discuss both *interested in* and BE *interested in*, two provide partial treatment (i.e. *interested* is contained in lists of 'Adjective+Preposition' patterns), and two do not treat it at all. As a result, none of the grammars or dictionaries presents all complementation patterns, either regarding *interested* in particular, or adjectives in general.

The grammars and dictionaries examined overlap to some extent in the complementation patterns they prioritise, background, or omit. Two complementation patterns are mostly neglected in both pedagogical grammars and dictionaries: *wh*-Clause and zero complementation. No grammar or dictionary treats '*interested in* + *wh*-Clause' explicitly: only two grammars present information about the pattern in a section on adjective complementation, and one contains the pattern in an exercise. Three dictionaries do not treat the pattern at all (they contain no information or examples), and two only provide examples. Zero complementation is more neglected in grammars: none treats the pattern explicitly, one presents it in a list of adjectives complementing a copular verb, adding that the adjectives may not have complements themselves, and one elicits zero complementation in an exercise (one sentence). In dictionaries, zero complementation is neglected to a lesser degree: although dictionaries do not treat it explicitly, they all include the pattern in at least one example.

The pedagogical grammars examined, taken collectively, also underrepresent the other three complementation patterns. The NP complementation is completely absent from two grammars, the others providing relevant information when discussing 'Preposition+Noun' patterns. In dictionaries, although NP complementation is not completely neglected, it is only treated explicitly and fully in two dictionaries, the other three only providing examples. The '*–ing* Clause' complement is absent from one grammar, the others providing relevant information when discussing 'Preposition+V-*ing*' patterns. Only one dictionary neglects this pattern, with three treating it explicitly and fully, and one providing examples. The *to*-inf complement is absent from one grammar and four have sections on 'Adj+*to*-inf'. Only one grammar provides explicit treatment of the *to*-inf pattern, and it also the only one that contrasts the '-*ing* Clause' and '*to*-inf' patterns. The '*to*-inf' pattern is the one receiving the most explicit attention in dictionaries, with four dictionaries explicitly mentioning the pattern (e.g. '*ADJ to*-inf'), and all five providing examples.

Finally, none of the pedagogical materials examined provides explicit information regarding alternative copular verbs, or make distinctions between use in written and spoken language. They rarely provide frequency information regarding complementation patterns, or

the likelihood of the copular verb being modalised in particular patterns – although some examples have modalised copulas. Also, no dictionary, and only one grammar, provide explicit information regarding the meaning of verbs that tend to be used in the '*to*-inf' complementation, and although all dictionaries but one provide relevant examples, only one grammar (the one offering the explicit treatment) does so. For an outline of the relevant content of each of the pedagogical materials examined here see the Appendix.

From the perspective of the usual users of pedagogical materials, neither grammars nor dictionaries (individually) provide all the information that a learner may need when seeking to use *interested* as an adjective (whether they are aware that it is an adjective or not), but the two types of pedagogical materials seem to complement one another to some extent. Another interesting observation is that they do not provide information only on the grammar or lexis part of the lexicogrammatical continuum, respectively. Pedagogical grammars approach language from the grammar end of the lexicogrammatical continuum (morphology and syntax), but they also provide some information on the grammatical patterns of particular words (e.g. *interesting* vs. *interested*). However, if a student wants to know the grammatical patterns that a word can be used in, then pedagogical grammars are of limited help. Dictionaries, on the other hand, approach language from the lexical end (foregrounding senses and lexical relations), but they also provide related information regarding grammatical patterns – either directly (in a sub-entry) or indirectly (in examples).

6.2 Corpus analysis

<u>Copular verbs complemented by the adjective *interested*</u>

The examination of the copular verbs complemented by the adjective *interested* revealed that BE is by far the most frequent copula complemented by *interested* in all four corpora, accounting for more than 92% of the instances in all corpus samples (Table 2). The examination of the overall frequencies of the copulas shown in Table 2 in BNCw and BNCs showed that the proportion of BE in this group of copulas is 84.4% and 79%, respectively (compared to 92.6% and 92.3% in the L1 corpus samples, respectively). That is, the predominance of BE in the pattern in focus cannot be fully explained by its very high overall frequency in the BNC. This seems to support the examination of *BE interested* as a pattern which is not only by far the most frequent, but also somehow distinct. The frequency comparisons of the other copulas in the samples also shows that there are clear differences between L1 and L2 use. Overall, L2 use contains a smaller proportion of other copular verbs – particularly in writing. In learner writing, *interested* complements copulas other than BE in just above one-tenth of the corresponding frequency in the BNCw sample, and the difference has high statistical significance (Table 3). Learner speaking contains other copulas in about one-third of the frequency in the spoken BNC, but the difference is not statistically significant (Table 4). More importantly, learners seem to use a much smaller variety of other copular verbs compared to native speakers: 4 types as opposed to 8 – with only BECOME, GET and MAKE being shared in the L1 and L2 corpus samples (Table 2).

Table 2. Copular verbs complemented by the adjective *interested*

Copular verbs	BNCw (N=215)	BNCw %	BNCs (N=234)	BNCs %	ICLE (N=236)	ICLE %	LINDSEI (N=210)	LINDSEI %
BE	199	92.6	216	92.3	234	99.2	198	94.3
Other	16	7.4	18	7.7	2	0.8	12	5.7
	BECOME (5) FEEL (1) GET (3) KEEP (2) MAKE (1) REMAIN (2) SEEM (2)		BECOME (4) GET (10) LOOK (1) MAKE (1) SEEM (2)		GET (1) MAKE (1)		BECOME (4) GET (6) MAKE (1) STAY (1)	

Table 3. Proportion of copular verbs other than BE: Comparison of ICLE and BNCw

	ICLE (N=236)	ICLE %	BNCw (N=215)	BNCw %	%DIFF	G2
Other copular verbs	2	0.9	16	7.4	722.2	**13.74**

Table 4. Proportion of copular verbs other than BE: Comparison of LINDSEI and BNCs

	LINDSEI (N=210)	LINDSEI %	BNCs (N=234)	BNCs %	%DIFF	G2
Other copular verbs	12	5.7	19	7.7	35.1	0.93

<u>Complementation patterns of *BE interested*</u>

In all four corpora, the most frequent complementation pattern by far is '*BE interested in +* NP' (Table 5). However, this is where the similarities stop, as there are differences in the proportions of all complementation patterns (even in the case of NP complementation) between L1 and L2 corpora, as well as between speech and writing. For example, in writing (both L1 and L2), the second most frequent pattern is '*BE interested in + ing*-Clause' (23.62% in BNCw, 11.5% in ICLE), whereas in speech, the second most frequent pattern is '*BE interested ∅*' in BNCs (30.23%), but '*BE interested in + ing*-Clause' in LINDSEI. The remainder of this section will compare the frequencies of the patterns as follows:
  a. speech and writing in L1
  b. speech and writing in L2
  c. writing in L1 and L2
  d. speech in L1 and L2

Table 5. *BE interested*: Proportion of complementation types in the four corpora

Complementation pattern	BNCw (N=199)	*BNCw %*	BNCs (N=216)	*BNCs %*	ICLE (N=234)	*ICLE %*	LINDSEI (N=198)	*LINDSEI %*
BE interested in + NP	102	*51.3*	90	*41.9*	185	*79.1*	149	*75.3*
BE interested in + *-ing* Clause	47	*23.6*	27	*12.5*	27	*11.5*	24	*12.1*
BE interested in + *wh*-Clause	7	*3.5*	13	*6.1*	11	*4.7*	8	*4.0*
BE interested + *to*-inf	11	*5.5*	21	*9.8*	4	*1.7*	5	*2.5*
BE interested ∅	32	*16.1*	65	*30.2*	7	*3.0*	12	*6.1*

Looking at L1 use (Table 6), the first observation is that the most frequent complementation pattern (*BE interested in* + NP) is also the only one with similar frequencies in both speech and writing. The only pattern with higher frequency in writing is '*BE interested in* + *-ing*-Clause', which shows almost double the frequency of the pattern in speech. The pattern that seems to characterise speech is zero complementation ('*BE interested* ∅'), which has almost double the frequency in speech compared to writing. The other two patterns ('*BE interested in* + *wh*-Clause' and '*BE interested* + *to*-inf') also have higher frequency (about 75%) in speech, but due to their low raw frequency in the corpora, the differences are not statistically significant.

Table 6. Complementation patterns of *BE interested*: Comparison of L1 speech and writing

Complementation pattern	BNCw (N=199)	*BNCw %*	BNCs (N=215)	*BNCs %*	%DIFF	G2
BE interested in + NP	102	*51.3*	90	*41.9*	22.5	1.97
BE interested in + *-ing* Clause	47	*23.6*	27	*12.5*	**88.1**	**7.13**
BE interested in + *wh*-Clause	7	*3.5*	13	*6.1*	**74.3**	1.39
BE interested + *to*-inf	11	*5.5*	21	*9.8*	**78.2**	2.45
BE interested ∅	32	*16.1*	65	*30.2*	**87.6**	**9.05**

As in L1, '*BE interested in* + NP' has very similar frequencies in L2 speech and writing (Table 7). However, the differences between speech and writing observed in the L1 corpora are not always mirrored in L2. The pattern '*BE interested* ∅' is twice as frequent in speech, compared to writing, in both L1 and L2 corpora, but in the case of L2 use the difference is not statistically significant, as raw frequencies in both L2 speech and writing are very low (see below for a discussion).

Table 7. Complementation patterns of *BE interested*: Comparison of L2 speech and writing

Complementation pattern	ICLE (N=234)	*ICLE %*	LINDSEI (N=198)	*LINDSEI %*	%DIFF	G2
BE interested in + NP	185	*79.1*	149	*75.3*	5.1	0.20
BE interested in + *-ing* Clause	27	*11.5*	24	*12.1*	5.2	0.03
BE interested in + *wh*-Clause	11	*4.7*	8	*4.0*	16.4	0.11
BE interested + *to*-inf	4	*1.7*	5	*2.5*	**47.1**	0.34
BE interested ∅	7	*3.0*	12	*6.1*	**103.3**	2.30

Table 8 summarises the results so far: the symbol '=' indicates similar frequencies in writing (W) and speech (S) (i.e. any differences are non-sizeable); '>' indicates frequency differences

that are both sizeable and statistically significant; parentheses indicate sizeable differences that are not statistically significant. Overall, L2 speech and writing show more similarities than differences, in contrast with the frequency differences in L1. In other words, L2 use seems more homogenised, in that it exhibits much less differentiation in the frequency of the five complementation patterns.

Table 8. Similarities and differences between L1 and L2 speech and writing

Complementation pattern	L1 corpora	L2 corpora
BE interested in + NP	W=S	W=S
BE interested in + *-ing* Clause	W>S	W=S
BE interested in + *wh-* Clause	(S>W)	W=S
BE interested + *to*-inf	(S>W)	(S>W)
BE interested ⌀	S>W	(S>W)

A more detailed picture of distributional differences emerges when we directly compare the frequencies of each complementation pattern in the written and spoken output of L1 and L2 users, respectively (Tables 9 and 10). In writing, four out of the five patterns show sizeable and statistically significant differences between L1 and L2. Learners tend to use three of the patterns less frequently than native speakers ('*BE interested in* + *ing*-Clause', '*BE interested in* + *to*-inf', and '*BE interested* ⌀') and one pattern more frequently ('*BE interested in* + NP'). Focusing on differences that are both sizeable and statistically significant, the comparisons in this section indicate a correlation between L2 use and the presence/absence and detail of the information on *(BE) interested* and its complementation patterns in pedagogical materials (particularly grammars).

The most prominent difference is in NP complementation, which is the only pattern that learners overuse in both speech and writing. In fact, not only do learners use it much more frequently than L1 users, but they also use it so frequently that it dominates their relevant output, as learners opt for NP complementation in more than three-quarters of the instances in the corpus samples. At first glance, this does not seem to correlate with the treatment of NP complementation in the pedagogical materials examined: the pattern may be fairly prominent in the dictionaries, but it is less than prominent in the grammars. However, grammars for lower levels (A1-A2) tend to only treat NP complementation (e.g. Murphy, 2007: 234-235). Therefore, it seems plausible to argue that NP complementation is rather neglected in grammars of higher levels (B1-C2), because the pattern is expected to be known. This is supported by the fact that two of the three B1-B2 grammars examined treat the NP pattern in sections discussing 'Adjective+Preposition' patterns, whereas only one of the three B2-C2 grammars mentions the pattern (without listing *interested*, and with no related examples or exercises). Also, the two B1-B2 grammars include the NP complementation in order to contrast it with the '–*ing* Clause' pattern, which is introduced at B1-B2 level.

Learners underuse two patterns in both speech and writing: zero complementation and '*to*-inf'. Zero complementation shows by far the largest difference: learners use the pattern about five times less frequently in speech and six times less frequently in writing. This correlates with the pattern's virtual absence from grammars, and the lack of explicit treatment in dictionaries (which only include the pattern in examples). The '*to*-inf' pattern is also clearly underused by learners (about four times less in speech and three times less in writing). This correlates with its treatment in pedagogical grammars, in which it is only discussed indirectly (as part of the 'Adjective + *to*-inf' pattern), but not with its treatment in dictionaries, as four out of five present the pattern explicitly. Finally, one pattern (–*ing*

Clause) is underused in L2 writing only, with learners having less than half of the proportion in L1 writing. Again, this correlates with the treatment of the pattern in grammars rather than dictionaries. However, the underuse of '*to*-inf' and '-*ing* Clause' patterns, and the predominance of NP complementation in L2 speech and writing, may also be attributed to syntactic complexity. As NP complementation is syntactically much less complex than the other two, learners may have opted for the simpler complementation pattern to avoid errors (e.g. Davydova, 2011; Vyatkina, 2013).

Table 9. Complementation patterns of *BE interested* in L1 and L2 writing

Complementation pattern	ICLE (N=234)	*ICLE %*	BNCw (N=199)	*BNCw %*	%DIFF	G2
BE interested in + NP	185	*79.1*	102	*51.3*	**54.2**	**12.78**
BE interested in + ing-Clause	27	*11.5*	47	*23.6*	**105.2**	**9.20**
BE interested in + wh-Clause	11	*4.7*	7	*3.5*	**33.6**	0.37
BE interested + to-inf	4	*1.7*	11	*5.5*	**223.5**	**4.63**
BE interested ⌀	7	*3.0*	32	*16.1*	**436.7**	**21.66**

Table 10. Complementation patterns of *BE interested* in L1 and L2 speech

Complementation pattern	LINDSEI (N=198)	*LINDSEI %*	BNCs (N=215)	*BNCs %*	%DIFF	G2
BE interested in + NP	149	*75.3*	90	*41.9*	**79.8**	**19.98**
BE interested in + ing-Clause	24	*12.1*	27	*12.5*	3.3	0.02
BE interested in + wh-Clause	8	*4.0*	13	*6.1*	52.5	0.83
BE interested + to-inf	5	*2.5*	21	*9.8*	**292.0**	**9.31**
BE interested ⌀	12	*6.1*	65	*30.2*	**395.1**	**35.87**

Modalisation of BE in *BE interested*

Looking at the complementation patterns of *BE interested* collectively in L1 use (Table 11), it becomes apparent that BE seems more likely to be modalised in speech than in writing, as modalised BE is about 70% more frequent in L1 speech, and the difference is statistically significant (G2=4.87). In terms of individual complementation patterns, the '*to*-inf' pattern contains by far the higher proportion of modalised BE in both L1 speech (57.1%) and writing (36.4%). In L1 writing, all other complementation patterns have very similar proportions of modalised BE (between 12.5% and 14.9%). However, there are distinct differences in L1 speech, with proportions of modalised BE ranging from 15.4% ('*wh*- Clause') to 33.3% ('-*ing* Clause'). The above indicate that both the medium (speech or writing) and the particular complementation pattern have a bearing on the likelihood that BE (in *BE interested*) will be modally marked.

In this light, it is important to observe two differences between L1 and L2 use. First, BE is modalised less frequently in L2, particularly in speech, where the difference is statistically significant. It occurs about three times less frequently in L2 speech (G2=15.69), and about a quarter less frequently in L2 writing (but G2=1.01). This, of course, may be due to the general tendency of learners to use modal marking less frequently than L1 users (e.g. Flowerdew, 2000; Gabrielatos & McEnery, 2005); however, the lack of any explicit information in pedagogical materials regarding modalisation of BE in *BE interested* might also be a contributing factor. Second, L2 use does not mirror the differences observed between L1

speech and writing: on average, the five complementation patterns, collectively, show almost identical proportions of modalised *BE interested* (10.6 in speech and 10.7 in writing). As in the case of complementation types, L2 use regarding modalisation of *BE interested* seems fairly homogenised. The remainder of the section will discuss frequency comparisons of modalised *BE interested* between L1 and L2 speech and writing in relation to the five complementation patterns.

Table 11. Proportion of modalised BE in the compelementation patterns of *BE interested*

	BNCw	*BNCw %*	BNCs	*BNCs %*	ICLE	*ICLE %*	LINDSEI	*LINDSEI %*
in+NP	13/102	*12.8*	15/90	*16.7*	22/185	*11.9*	11/149	*7.4*
in+ingC	7/47	*14.9*	9/27	*33.3*	2/27	*7.4*	3/24	*12.5*
in+whC	1/7	*14.29*	2/13	*15.4*	1/11	*9.1*	2/8	*25.0*
to-inf	4/11	***36.4***	12/21	***57.1***	1/4	***25.0***	0/5	***0***
⌀	4/32	*12.5*	14/65	*21.5*	0/7	*0*	1/12	*8.3*
Total	**29/199**	***14.6***	**52/216**	***24.7***	**26/234**	***11.1***	**17/198**	***8.6***

In L1, four out of the five complementation patterns have higher proportions of modalised BE in speech than in writing, although none of the differences is statistically significant (Table 12). L2 speech and writing also show differences, although not statistically significant (Table 13), but only in two instances do L1 and L2 show the same similarities/differences between speech and writing (*-ing* Clause, zero). That is, L2 use shows few similarities with L1 use regarding the frequency of modalised BE in different complementation patterns of *BE interested* in speech and writing

L1 and L2 writing show large differences in the proportions of modalised BE (Table 14): in four out of the five complementation patterns, learners have clearly lower proportions of modalised BE, but due to the small number of modalised instances, the differences are not statistically significant. However, there is an important relevant similarity: in both L1 and L2 writing, '*to*-inf' has by far the highest modal marking compared to other complementation patterns. In other words, the learners' frequency of modalised BE in the '*to*-inf' pattern tends to conform to L1 use (albeit to a smaller extent), despite the lack of explicit information in pedagogical materials. What can be tentatively hypothesised is that learners pick up useful cues from the modal marking in examples and exercises, which they replicate in their writing.

The situation is similar, but not identical, in L1 and L2 speech (Table 15). As in writing, learners modalise BE much less frequently in four out of the five complementation patterns. The only case when spoken L2 has higher modal marking than spoken L1 is '*wh*-Clause' complementation, but the difference is not statistically significant. However, in two patterns, NP and *to*-inf, the difference (lower proportions in L2) is so large that it is statistically significant despite the low frequencies involved. This may be explained by the almost identical proportions of modalisation in L2 speech and writing when complementation patterns are examined collectively (Table 11 above). As L1 shows much higher proportions in speech, the difference between L2 and L1 in speech is large enough to achieve statistical significance. Simply put, in the L2 corpora, speech tends to mirror writing.

Table 12. Proportion of modalised BE in the complementation patterns of *BE interested* in L1 speech and writing

	BNCw	*BNCw %*	BNCs	*BNCs %*	*%DIFF*	G2
in+NP	13/102	*12.8*	15/90	*16.7*	***30.5***	0.50
in+ingC	7/47	*14.9*	9/27	*33.3*	***123.5***	2.57
in+whC	1/7	*14.3*	2/13	*15.4*	*7.7*	0.00
to-inf	4/11	***36.4***	12/21	***57.1***	*56.7*	0.66
⌀	4/32	*12.5*	14/65	*21.5*	***72.0***	1.01

Table 13. Proportion of modalised BE in the compelementation patterns of *BE interested* in L2 speech and writing

	ICLE	*ICLE %*	LINDSEI	*LINDSEI %*	*%DIFF*	G2
in+NP	22/185	*11.9*	11/149	*7.4*	***61.1***	1.74
in+ingC	2/27	*7.4*	3/24	*12.5*	***68.8***	0.34
in+whC	1/11	*9.1*	2/8	*25.0*	*175.0*	0.73
to-inf	1/4	***25.0***	0/5	***0***	*2.5E+19*	1.62
⌀	0/7	*0*	1/12	*8.3*	***8.3E+18***	0.92

Table 14. Proportion of modalised BE in the complementation patterns of *BE interested* in L1 and L2 writing

	ICLE	*ICLE %*	BNCw	*BNCw %*	*%DIFF*	G2
in+NP	22/185	*11.9*	13/102	*12.8*	*7.6*	0.04
in+ingC	2/27	*7.4*	7/47	*14.9*	*101.4*	0.85
in+whC	1/11	*9.1*	1/7	*14.3*	*57.1*	0.10
to-inf	1/4	*25.0*	4/11	*36.4*	*45.6*	0.12
⌀	0/7	*0*	4/32	*12.5*	***1.25E+19***	1.58

Table 15. Proportion of modalised BE in the complementation patterns of *BE interested* in L1 and L2 speech

	LINDSEI	*LINDSEI %*	BNCs	*BNCs %*	*%DIFF*	G2
in+NP	11/149	*7.4*	15/90	***16.7***	*125.7*	**4.27**
in+ingC	3/24	*12.5*	9/27	*33.3*	*166.4*	2.47
in+whC	2/8	*25.0*	2/13	*15.4*	*62.6*	0.23
to-inf	0/5	***0***	12/21	***57.1***	*5.71E+19*	**5.13**
⌀	1/12	*8.3*	14/65	*21.5*	*159.0*	1.11

<u>Verb types in two complementation patterns of *BE interested*: *-ing* Clause and *to*-inf</u>

The analysis of collocates in the two L1 corpora pinpointed a distinct difference between the meaning of verbs in the two complement types. In the case of '*BE interested in* + *-ing* Clause', there is no particular meaning group that is more frequent than others; the verbs in the complement seem to be topic-specific. On the contrary, more than half of the verbs in '*BE*

interested + to-inf' have meanings that relate (directly or indirectly) to knowledge, or actions leading to knowledge (i.e. related to inquiry).

- Direct: *determine, discover, find out, know, learn, receive* (e.g. information), *share* (e.g. discovery), *study, understand*.
- Indirect: *analyse, assess, check, compare, contrast, discuss, examine, experience, explore, hear, identify, interview, listen, look, monitor, notice, observe, read, research, see, speak, study, talk, test, visit, watch, witness*.

Tables 16 and 17 show the proportions of knowledge-related verbs in the two complementation patterns (proportions are calculated on all collocates with MI≥3).

Table 16. Proportion of knowledge-related verbs in the two complementation patterns: BNCw

BNCw	Knowledge	All	%
BE interested + to-inf	22	41	*53.7*
BE interested in + -ing Clause	23	301	*7.6*

Table 17. Proportion of knowledge-related verbs in the two complementation patterns: BNCs

BNCs	Knowledge	All	%
BE interested + to-inf	8	14	*57.1*
BE interested in + -ing Clause	7	48	*14.6*

The proportion of knowledge-related verbs in '*BE interested + to*-inf', as compared to the proportion in '*BE interested in + -ing* Clause') is seven times higher in L1 writing and four times higher in L1 speech, and both differences are statistically significant despite the very low frequencies (G2=36.85 and G2=6.66 respectively). This suggests that the verbs in the '*to*-inf' complement strongly tend to be pattern-specific. However, the preference for knowledge-related verbs is probabilistic rather than absolute. It is not that the '*–ing* Clause' complement does not contain knowledge-related verbs; rather, it seems much more likely for such verbs to be used in the '*to*-inf' complement. Before moving on to the comparisons with L2 speech and writing, we need to discuss the theoretical implications of these findings, particularly regarding the treatment of collocation in Lexical Grammar as "a purely lexical relation, non-directional and probabilistic, which ignores any syntactic relation between the words" (Stubbs, 2001: 64). The results contradict this treatment, as the adjective *interested* has distinctly different collocation patterns when the collocates are in different grammatical patterns (in our case, '*BE interested in + -ing* Clause' and '*BE interested + to*-inf'). The tendency to use knowledge-related verbs in '*to*-inf' complementation is also present in L2 use (Tables 18 and 19). However, differences are much less distinct (about 2.5 times more frequent in both writing and speech), and due to the extremely low frequencies they are not statistically significant (G2=1.42 and G2=0.58, respectively).

Table 18. Proportion of knowledge-related verbs in the two complementation patterns: ICLE

BNCw	Knowledge	All	%
BE interested + to-inf	3	5	*60.0*
BE interested in + -ing Clause	10	39	*25.6*

Table 19. Proportion of knowledge-related verbs in the two complementation patterns:
LINDSEI

BNCs	Knowledge	All	*%*
BE interested + *to*-inf	1	1	*100*
BE interested in + *-ing* Clause	6	15	*40.0*

Looking at the results from a different angle, learners seem to use knowledge-related verbs in '*-ing* Clause' complements much more frequently than L1 users (about 3.5 times more in writing and about three times more in speech). In other words, when learners want to use a knowledge-related verb as a complement of the adjective *interested*, they tend to opt much less frequently for the complementation pattern preferred by L1 users.

7. CONCLUSIONS, SUGGESTIONS & FURTHER STEPS

The corpus analysis of L1 use revealed lexicogrammatical patterns that are not included in either reference or pedagogical materials, and showed not only that are there frequency differences between the complementation patterns of BE *interested*, but also that these patterns have different frequencies in speech and writing. Also, both the process and results of the analysis support the conception of lexis and grammar as intertwined, without, however, providing support for treating either as primary. Although the adjective *interested* can complement a number of copular verbs, the analysis revealed that BE is by far the most frequent one. Also, some complementation patterns show distinctly higher frequency of modalised BE, and others show a clear preference for verbs of a particular meaning in the complement. The latter indicates that a collocation analysis of the word form *interested* (particularly in the usual 5L-5R span) would mainly return collocates of *interested* in its most frequent word class, and in the most frequent syntactic patterns the word is found. This type of analysis would be useful for the examination of the general semantic associations of a word-form (a common technique in corpus-based critical discourse studies – e.g. Partington, Duguid & Taylor, 2013), but it would not contribute to lexicogrammatical description or theory. That is, it would not reveal that '*BE interested* + *to*-inf' shows preference for a distinctive group of verb collocates in its complement, whereas '*BE interested in* + *-ing* Clause' does not. In fact, this grammar-specific collocation distinctiveness would not be observed even if the node was *interested* as an adjective, or if collocates were restricted to verbs, as it would fail to distinguish between co-occurrences within different syntactical structures. Whatever the focus and starting point of the analysis, all patterns (and their components) can only be fully defined if both lexical and grammatical aspects are taken into account. We can, of course, focus on lexis to get lexis-like answers, but this should not mislead us to see grammar as secondary, or conclude that any patterns observed in such an approach can be explained in terms of lexis only, or that the lexis we choose to focus on (our starting point) is at the core of the patterns we observed. A related observation is that Halliday's (1992: 64) "tunnelling" metaphor may not be entirely useful, as it seems to imply both linearity and directionality in research: from the lexis end towards the grammar end of the continuum (or vice versa). Rather, the present analysis suggests that lexicogrammatical research cannot be mono-directional: at any given point of the analysis both grammar and lexis are involved. This observation is also related to the flexibility in the choice of the starting point of the analysis. For example, in Lexical Grammar the starting point is always a word/phrase, whereas in collostructional analysis the starting point is always a complex construction. In the same vein, whether the derived answer is perceived as lexis-like or grammar-like can be a matter of perspective or theoretical orientation. For example, a study

may examine the frequency that a particular verb, or a semantically-defined group of verbs, is used in the progressive aspect, whereas another study may examine the frequency that the progressive aspect is used with particular verbs. Despite their different starting points, both studies would be essentially examining the same lexicogrammatical item.

Comparisons of L2 and L1 use revealed more differences than similarities. Overall, the results suggest a correlation between learner use and the treatment of *BE interested* in pedagogical materials, that is, the extent and detail of information they provide, and the relevant examples and exercises. Of course, it cannot be claimed that the correlations identified here necessarily point towards a direct influence of the content of pedagogical materials on learner use. However, these correlations are numerous and sizeable enough to merit consideration in decisions regarding the relevant content of pedagogical materials. More precisely, pedagogical materials at B1-C2 levels could usefully provide more comprehensive and nuanced information on the complementation and related lexicogrammatical patterns of *BE interested* (and, more generally, 'copula+*interested*'), and this information should be informed by the frequency of these patterns in speech and writing. In light of the differences in the focus and coverage of the two types of pedagogical materials examined here, L2 use seems to correlate with the information in grammars rather than dictionaries. This may be because lexicogrammatical patterns such as the ones examined in this study are presented to learners as grammatical points and/or that dictionaries are treated as sources for the meaning of words rather than their use (perhaps excepting collocation). Interestingly, the comparison between the grammars and dictionaries examined in this study (taken collectively) suggests that, in the case of lexicogrammatical patterns such as 'copula+*interested*', dictionaries would seem a more useful source for learners. Having said that, the overarching conclusion that can be drawn regarding pedagogical materials is that there is no single source which provides all the lexicogrammatical information that a learner may need in order to form a comprehensive picture of the use of 'copula+*interested*'. However, the analysis also indicated that some sources provide more information than others, which suggests that the selection of sources (by teachers or learners) is crucial. This also strongly indicates that consulting a single pedagogical source cannot be expected to be sufficient, and learners would be wise to combine sources, in particular grammars and dictionaries.

The analysis of pedagogical materials also showed that grammars and dictionaries overlap in their coverage: grammars also provide lexis-like information, and dictionaries also provide grammar-like information. However, the very existence of grammars and dictionaries for learners indicates that language teaching still treats grammar and lexis in a fairly compartmentalised fashion. When thinking of hard-copy publications, the distinction makes practical sense (as there are size and cost limitations); however, online publishing offers possibilities for more comprehensive learner resources that combine the features of pedagogical grammars and dictionaries: *pedagogical lexicogrammars*. Such resources would not only be easily updatable and expandable, but their content would also be interlinked. Using such a resource, learners would be able to access language information starting at any point of the lexicogrammar continuum, and then move back and forth along the continuum combining the information they access. For example, a learner looking up *interested* to find information on how to use it in a sentence (or wishing to check if their sentence contains a correct use of *interested*), that is, a learner starting at the lexis-end of the continuum, would not only encounter dictionary-like information, but also links to adjective complementation patterns, and complementation in general (the grammar end), as well as information on the types of verbs specific to particular patterns (more towards the lexis end), and issues of modal marking (around the middle). Finally, it would be feasible for entries to have links to corpora, so that learners could examine relevant concordance lines, a practice that, apart from

providing a wealth of examples of actual use, would also allow for "serendipity" (Bernardini, 2000), that is opportunities for learners to discover language features other than the ones for which they accessed the pedagogical lexicogrammar.

The present study has not provided a full lexicogrammatical profile of the adjective *interested*. Further research needs to examine the complementation patterns of *interested* when it complements other copulas. The analysis of the modal marking needs to examine all copulas, and must be expanded to looking at the distribution of different modality types. In addition, an analysis of the type of Subject of 'copula+*interested*' (e.g. noun, pronoun, impersonal *it*) can reveal correlations between type of Subject and particular complementation patterns. Finally, it would be useful to examine whether these patterns are specific to *interested*, or whether the type of adjective (e.g. in terms of meaning) influences lexicogrammatical patterns. Of course, all the above lexicogrammatical patterns should also be examined in different varieties of English, and different genres within the same variety, as well as written and spoken L1 corpora comparable to ICLE and LINDSEI in terms of genre and user age (e.g. LOCNESS and LOCNEC).

References

Bernardini, S. (2000) Systematising serendipity: Proposals for concordancing large corpora with language learners. In Burnard, L. & McEnery, T. (eds), *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the third international conference on teaching and language corpora*. (Lodz Studies in Language). Hamburg: Peter Lang. 183-190.

Biber, D. & Reppen, R. (2002) What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition,* 24(2), 199-208.

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

Biber, D., Conrad, S., & Reppen, R. (1998) *Corpus Linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan E. (1999) *Longman Grammar of Spoken and Written English*. Harlow, Essex: Longman.

Carter, R. & McCarthy, M. (2006) *Cambridge Grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.

Culicover, P.W., Jackendoff, R. & Audring, J. (2017) Multiword constructions in the grammar. *Topics in Cognitive Science*. Early View. 1-17. doi:10.1111/tops.12255.

Davydova, J. (2011) *The Present Perfect in non-native Englishes: A corpus-based study of variation*. Berlin: Walter de Gruyter.

Deshors, S.C. (2015) A multifactorial approach to gerundial and to-infinitival verb-complementation patterns in native and non-native English. *English Text Construction*, 8(2), 207-235.

Ellis, N.C. (2015) Implicit AND explicit learning: Their dynamic interface and complexity. In Rebuschat, P. (ed.) *Implicit and Explicit Learning of Languages* (pp. 3-23). Amsterdam: John Benjamins.

Ellis, R. (2015) *Understanding Second Language Acquisition* (2nd ed.) Oxford: Oxford University Press.

Erman, B. & Warren, B. (2000) The idiom principle and the open choice principle. *Text*, 20(1), 29-62.

Flowerdew, J. (2000) Investigating referential and pragmatic errors in a learner corpus. In Burnard, L. & McEnery, T. (eds.), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang. 145-154.

Francis, G. (1993) A corpus-driven approach to grammar: Principles, methods and examples. In Baker, M. Francis, G. & Tognini-Bonelli, E. (eds.) *Text and Technology. In Honour of John Sinclair*. Amsterdam: Benjamins. 137-156.

Gabrielatos, C. (2003). Conditional sentences: ELT typology and corpus evidence. *36th Annual BAAL Meeting*, University of Leeds, UK, 4-6 September 2003.

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ* 8(4), 1-37, http://www.tesl-ej.org/wordpress/issues/volume8/ej32/ej32a1.

Gabrielatos, C. (2006) Corpus-based evaluation of pedagogical materials: *If*-conditionals in ELT coursebooks and the BNC. *7th Teaching and Language Corpora Conference*, University Paris 7 Denis Diderot, Paris, France, 1-4 July 2006.

Gabrielatos, C. (2013) *If*-conditionals in ICLE and the BNC: A success story for teaching or learning? In Granger, S., Gilquin, G. & Meunier, F. (eds.) *Twenty Years of Learner Corpus Research: Looking back, moving ahead*. Louvain-la-Neuve, Belgium: Presses Universitaires de Louvain. 155-166.

Gabrielatos, C. & McEnery, T. (2005) Epistemic modality in MA dissertations. In Fuertes Olivera, P.A. (ed.), *Lengua y Sociedad: Investigaciones recientes enlingüística aplicada. Lingüística y Filología no. 61*. Valladolid: Universidad de Valladolid. 311-331.

Gilquin, G. (2012) The non-finite verb slot in English causative constructions: Comparing native and learner collostructions. In Leino, J. & von Waldenfels, R. (eds.) *Analytical Causatives: From "give" and "come" to "let" and "make"*. München: Lincom Europa.

Gilquin, G., De Cock, S., & Granger, S. (eds.) (2010) *LINDSEI: Louvain International Database of Spoken English Interlanguage*. Louvain, Belgium: UCL Presses.

Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. (2009) *International Corpus of Learner English*. Presses universitaires de Louvain, Louvain-la-Neuve.

Granger, S., Gilquin, G. & Meunier, F. (eds.) (2013) *Twenty Years of Learner Corpus Research: Looking back, moving ahead*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Granger, S., Gilquin, G. & Meunier, F. (eds.) (2015) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

Gries, S.Th. & Stefanowitsch, A. (2004) Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics,* 9(1), 97–129.

Halliday, M.A.K. (1961) Categories of a theory of grammar. *Word*, 17(3), 241–292.

Halliday, M.A.K. (1966) Lexis as a linguistic level. In Bazell, C.E., Catford, J.C., Halliday, M.A.K. & Robins, R.H. (eds.) *In memory of F. R. Firth*. London: Longman. 148–62.

Halliday, M.A.K. (1991) Corpus studies and probabilistic grammar. In Aijmer, K. & Altenberg, B. (eds.), *English Corpus Linguistics: Studies in honour of Jan Svartvik*. London: Longman. 30-40.

Halliday, M.A.K. (1992) Language as system and language as instance: the corpus as a theoretical construct. In Svartvik, J. (ed.), *Directions in corpus linguistics: proceedings of the Nobel Symposium 82, Stockholm, 4–8 August 1991*. Berlin: Mouton de Gruyter. 61–77.

Hardie, A. (2012) CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics,* 17: 380-409.

Harwood, N. (2005) What do we want EAP teaching materials for? *Journal of English for Academic Purposes* 4(2), 149-161.

Hasselgård, H. (2016) *The way of the world:* The colligational framework "the N1 of the N2" and its Norwegian correspondences. *Nordic Journal of English Studies*, 15(3), 55-79.

Hasselgren, A. (1994) Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4, 237–258.

Hoffmann, S., Evert, S., Smith, N., Lee, D. & Berglund-Prytz, Y. (2008) *Corpus Linguistics with BNCweb - a practical guide.* Frankfurt am Main: Peter Lang.

Hoey, M. (1997) From concordance to text structure: New uses for computer corpora. In Melia, J. & Lewandoska, B. (eds.), *PALC '97: Practical Applications in Language Corpora*. Łódź: Łódź University Press. 2-23.

Hoey, M. (2005) *Lexical Priming: A new theory of words and language*. London, England: Routledge.

Huddleston, R. & Pullum, G.K. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hunston, S. & Francis, G. (1998) Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics,* 19(1), 45-72.

Hunston, S. & Francis, G. (2000) *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam, Netherlands: John Benjamins.

Kennedy, G. (1992) Preferred ways of putting things with implications for language teaching. In Svartvik, J. (ed.) *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82*, Stockholm, 4-8 August 1991. Berlin: Mouton de Gruyter. 335-378.

Leech, G. (2011) Frequency, corpora and language learning. In Meunier, F., De Cock, S., Gilquin, G. & Paquot, M. (eds.) *A Taste for Corpora: In honour of Sylviane Granger*. Amsterdam: John Benjamins. 7-31.

McEnery, A. & Xiao, Z. (2005) HELP or HELP to: What do corpora have to say? *English Studies*, 86(2), 161-187.

Meunier, F. (2012) Formulaic language and language teaching. *Annual Review of Applied Linguistics* 32(1), 111-129.

Meunier, F. & Gouverneur, C. (2009) New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. In Aijmer K. (ed.) *Corpora and Language Teaching*. Amsterdam and Philadelphia: Benjamins. 179-201.

Ortega, L. (2013) *Understanding Second Language Acquisition*. Oxford: Routledge

Owen, C. (1993) Corpus-based grammar and the Heineken effect: Lexicogrammatical description for language learners. *Applied Linguistics* 14(2), 167-187.

Partington, A., Duguid, A. & Taylor. C. (2013). *Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam & Philadelphia: John Benjamins.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Römer, U. (2004) Textbooks: A corpus-driven approach to modal auxiliaries and their didactics. In Sinclair, J.McH (ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 185–199.

Römer, U. (2009) The inseparability of lexis and grammar. *Annual Review of Cognitive Linguistics* 7, 140–162.

Rühlemann, C. (2007) Lexical grammar: The GET-passive as a case in point. *ICAME Journal*, 31, 111-128.

Sinclair, J. (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Sinclair J. (1996) The search for units of meaning. *Textus,* 9(1), 75-106.

Sinclair, J. (2004) *Trust the Text: Language, corpus and discourse*. London: Routledge.

Stefanowitsch, A. & Gries, S.Th. (2003) 'Collostructions: Investigating the interaction of words and constructions.' *International Journal of Corpus Linguistics* 8 (2), 209-243.

Stubbs, M. (1996) *Text and Corpus Analysis: Computer-assisted studies of language and culture.* Oxford: Blackwell.

Stubbs, M. (2002) Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics,* 7(2), 215-244.

Stubbs, M. (2001) *Words and Phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Stubbs, M. (2009) The search for units of meaning: Sinclair on empirical Semantics. *Applied Linguistics,* 30(1), 115–137.

Tomlinson, B. (2013) Introduction: Applied linguistics and materials development. In Tomlinson, B. (ed.) *Applied Linguistics and Materials Development*. London: Bloomsbury. 1-7.

Vyatkina, N. (2013) Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *Modern Language Journal* 97, 11-30.

Xu, J. & Wu, L. (2014) Web-based fourth generation corpus analysis tools and the BFSU CQPweb case. *Waiyu Dianhua Jiaoxue* [Computer-assisted Foreing Language Education]). http://en.cnki.com.cn/Article_en/CJFDTotal-WYDH201405002.htm

Pedagogical materials examined in this study

Bunting, J.D., Diniz, L. & Reppen, R. (2012). *Grammar and Beyond Level 4, Student's Book* Cambridge: Cambridge University Press.

Cambridge Dictionary Online. http://dictionary.cambridge.org

Collins COBUILD English Grammar.

Collins Dictionary Online. collinsdictionary.com

Davis, F. & Rimmer, W. (2011). *Active Grammar Level 2*. Cambridge: Cambridge University Press.

Hewing, M. (2013) *Advanced Grammar in Use* (3rd ed.). Cambridge: Cambridge University Press.

Lloyd, M. & Day, J. (2011). *Active Grammar Level 3*. Cambridge: Cambridge University Press.

Longman Dictionary of Contemporary English Online: http://ldoceonline.com

Macmillan Dictionary Online: http://macmillandictionary.com

Murphy, R. (2007) *Essential Grammar in Use* (3rd ed.). Cambridge: Cambridge University Press.

Murphy, R. (2012). *English Grammar in Use* (4th ed.). Cambridge: Cambridge University Press.

Oxford Advanced Learner's Dictionary Online: http://oxfordlearnersdictionaries.com