

I. ARTICLES

GRAHAM BURTON

Facoltà di Scienze della Formazione, Libera Università di Bolzano

'Are you going to go?' **Putting a pedagogical grammar rule** **under the corpus spotlight**

ABSTRACT. A rule stating that we tend to avoid using *go* and *come* after the future marker *going to* appears again and again in many coursebooks and grammars used in English Language Teaching, and has done for decades. This article attempts to show, using empirical evidence from corpora, why the rule is inaccurate, and different ways that this might be established. As the rule under consideration is typically framed as a tendency (like many other pedagogical grammar rules), an additional aim of the work is to outline the kinds of corpus analyses researchers and materials designers can potentially use in order to investigate the question of (claimed) linguistic tendencies. The article concludes by discussing why a rule that is apparently inaccurate nevertheless appears again and again in print, arguing that the existence of a well-established and widely-accepted 'canon' of ELT grammar means that such inaccuracies in descriptions of grammar can be easily perpetuated.

KEYWORDS: English Language Teaching, pedagogical grammar, coursebooks, materials design, corpus linguistics.

1. INTRODUCTION

When learners of English as a Foreign Language (EFL) study the future form *going to*, they are typically taught that it is used to express two future functions: talking about plans and intentions, and making predictions when there is present evidence. For example, Raymond Murphy's *English grammar in use* (Murphy 2019: 40) defines these two uses as follows:

I am going to do something = I have already decided to do it, I intend to do it. [...]
When we say that 'something is going to happen', the situation now makes this clear.

However, in many pedagogical grammars and coursebooks, an extra rule is often given. It is a rule that is rarely given a great deal of prominence (in the case of coursebooks, it is often included in a grammar reference section at the back of the book, rather than in the main lesson sections). Nevertheless, it appears repeatedly, in many different publications, and, as will be seen later in this article, has been appearing repeatedly for nearly a century. The rule in question is that which states that we do not usually use the verb *go* after the future marker *going to*. This rule, or some variation of it, can be found in both coursebooks and practice grammars.

The aim of this article is twofold. Firstly, it attempts to assess the veracity of the rule, by comparing it to corpus evidence, in order to ascertain whether or not it should be taught to learners of EFL. However, as will be seen, the fact that the rule is typically phrased as a tendency brings complications to this kind of comparison, and the second aim of this article is therefore to explore how a claimed linguistic 'tendency' might be confirmed, or refuted, through corpus evidence.

2. LITERATURE REVIEW

2.1. *Going to go* in English Language Teaching descriptions

In materials used in English Language Teaching (ELT), the rule outlawing *going to go* is diffuse, although not ubiquitous, and can be found in a variety of publications. It appears, for example, in three of the most popular General (British) English coursebook series for adults: *Headway* (Oxford University Press), *New English file* (Oxford University Press) and *Face2face* (Cambridge University Press). It can also be found in a number of recent pedagogical grammars, including *MyGrammarLab advanced* (Foley & Hall 2012a) and *Oxford practice grammar* (Eastwood 2003). It also features in a pedagogical grammar aimed at teachers of EFL, *A concise grammar for English language teachers* (Penston 2009). On the other hand, the rule is absent in three of the most influential pedagogical grammars in the last few decades: *Murphy's English grammar in use* (2019), *Azar and Hagen's Fundamentals of English grammar* (2011) and *Swan's Practical English usage* (2016).

Looking at the rule in more detail, it becomes apparent that there are actually three versions. The first, and most limited version, states that the restriction after *going to* applies only to the verb *go*. The second version says it applies not only to *go* but also to *come*. Finally, the strongest version states that it applies to 'verbs of movement' in general. Table 1 summarises the three different versions of the rule, with an example for each version from published ELT materials.

Table 1. The three versions of the *going to* rule

| Restriction | Examples |
|---|--|
| <i>go</i> after <i>going to</i> | "With the verb <i>go</i> , we usually say, <i>I'm going to Italy</i> , not <i>I'm going to go to Italy</i> , but both are correct." (Redston & Cunningham 2005: 129) |
| <i>go</i> and <i>come</i> after <i>going to</i> | "We usually avoid <i>be going to</i> with the verbs <i>go</i> and <i>come</i> ." (Foley & Hall 2012a: 124) |
| verbs of movement after <i>going to</i> | "With verbs of movement, especially <i>go</i> and <i>come</i> , we often use the present continuous rather than <i>going to</i> ." (Eastwood 1994: 97) |

One important characteristic of the way that the *going to* rule is expressed is that it is invariably described in relative rather than absolute terms; the claim is that speakers 'tend not to' produce or 'avoid' producing forms like *going to go* or *going to come*, not that speakers never produce them. This kind of wording is not unusual in pedagogical grammar rules, and it is easy to find similar phrasing in rules relating to other areas of grammar. The following are all examples of similarly worded rules, taken from the titles mentioned at the beginning of this section:

"We don't usually use *going to* in short answers: *Yes, she is*. NOT *Yes, she's going to*." (*Face2face pre-intermediate student's book*, Redston & Cunningham 2005)

"We usually use the past simple, not the past perfect if the series of actions is clear." (*MyGrammarLab: intermediate*, Foley & Hall 2012b)

"We don't usually use activity / state verbs in continuous verb forms." (*Face2face upper intermediate student's book*, Redston & Cunningham 2016)

"We don't usually say *a bike / bag of hers* (unstressed), etc, but we do say *a friend / daughter / student / play of hers*, etc, showing that what is possessed is usually a (indefinite) person or creative work." (*A concise grammar for English language teachers*, Penston 2009)

"We do not usually say *a milk* or *two soups*." (*Oxford practice grammar*, Eastwood 2003)

It is perhaps understandable why pedagogical grammar rules for learners of EFL are sometimes framed in this way. Firstly, stating that we 'don't usually / normally say X' is essentially a frequency-based argument: it is equivalent to saying 'It is infrequent to say X', with the implication that since native speakers say X infrequently, then learners also ought not to say it frequently, because to do so would be a mistake. Frequency-based arguments in grammar descriptions are well established in ELT pedagogical grammar, going back to at least the first half of the twentieth century; for example, W.S. Allen, in his hugely successful grammar *Living English structure* (1947: vii), speaks of using 'personal structure counts' in its construction and this has continued into the present (see, for example, Biber & Reppen's [2002] frequency-based arguments on how grammar

syllabuses for EFL should be revised in order to reflect usage, as shown by corpus-based frequency studies).

It is also possible to see such wordings as a form of ‘hedge’, that is to say, the use of a linguistic device which allows an author to display a “lack of commitment to the truth of a statement or a desire not to express that commitment categorically” (Richards & Schmidt 2007: 237); in this case, it is the truth of a pedagogical grammar rule to which a materials designer, choosing to word the rule in this way, attempts to withhold complete commitment. The idea that a grammarian might provide a grammar rule that she or he does not believe to be entirely true is not controversial; the ELT grammarian Michael Swan, in discussing ‘design criteria’ for pedagogical grammar rules, states that while it is “desirable to tell learners the truth [...], one will often need to compromise with truth for the sake of clarity, simplicity, conceptual parsimony or relevance” (Swan 1994: 46).

2.2. *Going to* in general grammatical descriptions

The situation in more general grammatical descriptions of English is markedly different. There is no reference to restrictions on verb choice after *going to* in major academic grammars such as *The Cambridge grammar of the English language* (Huddleston & Pullum 2002), *The Cambridge grammar of English* (Carter & McCarthy 2006) and *A comprehensive grammar of the English language* (Quirk, Leech & Greenbaum 1985).

One relatively recent exception is Declerck, Reed and Cappelle (2006) *The grammar of the English tense system: A comprehensive analysis*, which states that “the use of *be going to* sounds rather awkward before *go* and *come*” (2006: 355). Again, the analysis is worded as a tendency, with the authors claiming that “[p]ople therefore tend to use the present progressive form of these verbs instead”. However, later in the very same title the example sentence ‘I was going to go to Cuba this summer’ (2006: 581) is given to illustrate the discussion of a completely different area of grammar, without any reference to awkwardness in the combination of *going to* and *go*.

2.3. Historical references

As shown in the previous sections, the rule outlawing *going to* before *go*, *come* and verbs of movement is often found in contemporary ELT accounts, but not typically found in general grammars. This section will consider the historical

aspect; what, if anything, do older grammars and older ELT materials have to say about restrictions on verbs after *going to*?

A search of the HUGE database – a searchable corpus of English grammars and usage guides for the period 1770–2010 (see huge.ullet.net) – produces only one hit for the phrase *going to go*, from Fitzgerald's (1901) *Word and phrase*. In this grammar, the phrase is mentioned disapprovingly but only in passing, along with the use of *going to* more generally, in a section sub-titled "A group of slovenly phrases":

'Going to go' is plainly inelegant and almost fatuous, however idiomatic and customary it may be; but the use of "I am going to" do this or that, as a habitual mode of expressing purpose to perform an action, is almost as objectionable (Fitzgerald 1901: 333).

That only one title in the HUGE database mentions *going to go* at all shows that little attention has been paid historically to verb choice after *going to* and that the restrictions reported in ELT accounts do not originate in the famously prescriptive older grammars of English (see Locher [2008] for an account of how prescriptivism became rife in English grammars from the 18th century onwards). Further, Fitzgerald's decision to mention *going to go* at all, and the wording of his comment "however idiomatic and customary it may be", very strongly imply that the phrase was widely used, and not avoided at all, despite his personal dislike of it.

The wave of descriptive (sometimes also known as 'scientific') grammars of English that appeared around the turn of the 20th century also have nothing to say about phrases such as *going to go*. A rule stating that it is avoided or should not be used cannot be found in well-known titles such as Nesfield's (1898) *English grammar past and present*, Sweet's (1898) *A new English grammar, logical and historical*, and Kruisinga's (1931) *A handbook of present-day English*. Interestingly, Jespersen's (1933) *Essentials of English grammar* actually refers to the fact that it is possible to say *going to come* and *going to go*, in order to make the point (illustrated with examples from Dickens) that "*going* has become a mere grammatical implement and no longer implies any movement" (Jespersen 1933: 268).

The situation seems to have changed with the publication of Harold Palmer's (1924) *A grammar of spoken English on a strictly phonetic basis*. This was a highly innovative grammar, intended to be used chiefly (but not exclusively) by foreign adult students of English" (Palmer 1924: xxx), and is explicit in its proscription: the verbs *come* and *go*, states Palmer, are "generally excluded from this category", in other words, excluded from the category of verbs used after *going to* (Palmer 1924: 280). Subsequent to this, versions of the rule can be found in widely used ELT pedagogical grammars such as *A practical English grammar* (Thomson &

Martinet 1960/1969), and *Longman English grammar* (Alexander 1998). On the other hand, the rule is missing from W. Stannard Allen's (1947) *Living English structure*, a hugely successful early pedagogical grammar.

In short, an analysis of historical and contemporary primary literature suggests that the rule outlawing *go, come* and other verbs of movement is a relatively recent development, appearing only in the early 20th century. Furthermore, it is almost exclusively associated with accounts provided in ELT materials, and is rarely found in more general accounts of English grammar.

2.4. Corpus-based critiques

As stated, one of the aims of this paper is to provide a critique of the pedagogical grammar rule under examination, with reference to corpus evidence. As Swan (1994) has argued, pedagogical grammar rules are born of a series of often conflicting criteria, and from this point of view, it is therefore perhaps unsurprising that a number of studies have called into question the accuracy of many aspects of grammatical accounts of EFL that appear in print.

There have been few studies dealing specifically with the coverage of future forms in ELT grammar explanations, and none to specifically focus on the *going to* rule. More general examinations of the semantics and pragmatics of *going to* have been carried out, e.g. Brisard (2001). However, Close (1970a, b) reviews the coverage of future forms in older grammars and suggests a possible four-way pedagogical classification system for futurity. Tregidgo (1980) discusses the uses of *will* and *shall* in contemporary English, recommending teaching points up to Intermediate levels; and Locke (1986) discusses three different uses of the future continuous, criticising the coverage of the form in many ELT accounts.

More generally, however, tense and aspect have constituted an area for corpus-based investigation of published ELT materials. For example, Shortall's (2007) study focussed on treatments of the present perfect in coursebooks, comparing them to corpus evidence. Römer (2005) compared coverage of simple and progressive forms in coursebooks with corpus evidence, and Vandenhoeck (2018) examined the representation of the past perfect in coursebooks, again comparing this to evidence from corpora. Other examples of studies to have compared corpus evidence to the presentation of grammar in teaching materials for EFL include those focussing on conditional structures (Frazier 2003; Gabrielatos 2003, 2006; Jones & Waller 2011), reported speech (Barbieri & Eckhardt 2007) and the adverb *though* (Conrad 2004).

The conclusions of such studies are typically that the teaching materials examined misrepresent actual usage identified through corpus analysis; nonetheless,

the grammatical descriptions in coursebooks and pedagogical grammars have remained unchanged. Possible explanations offered for this have related to questions of whether corpus data taken outside its original context is still authentic (Prodromou 2003), or whether non-corpus data might be favoured over corpus data for pedagogical reasons (Carter 1998; Cook 1998; Shortall 2007; Timmis 2015). On the other hand, Burton (2012) found that many coursebook writers simply do not have the time or have not had the training to use corpora, and, in some cases, have little interest in or motivation to use them. The commercial aspect is also likely to be significant; publishers around the world are likely to be hesitant about changing or updating grammatical descriptions, fearing that teachers can be easily alienated (Burton 2012; Littlejohn 1992). We will return to this point later in this article.

3. RESEARCH QUESTIONS

The main aim of this article is to assess, by reference to corpus evidence, the validity of the oft-repeated statement, found in many pedagogical grammar treatments for EFL, that we tend to avoid using the verbs *go*, *come* or, more generally, verbs of movement, after the future marker *going to*. Only the behaviour of the verbs *go* and *come* is considered; since these are both 'verbs of movement', any observations made about their behaviour can also logically be applied to the broader version of the rule that includes verbs of movement.

As we have seen, the rule under examination is typically hedged – that is, stated as a tendency. A rule simply stating that the verbs *go* and *come* cannot be used after *going to* would be relatively easy to refute using frequency evidence from corpora; the only difficulty would be in establishing the minimum number of instances of *going to go* or *going to come* in the corpus that could be considered sufficient evidence to disprove it. However, when a rule is expressed as a tendency, the situation becomes more complicated, since it would always predict or allow a certain number of attested incidences without automatically being considered false. This paper will therefore also address the question of which kinds of corpus tests can be employed in order to investigate a (claimed) linguistic tendency. The two research questions under investigation are therefore as follows:

RQ1: Does corpus evidence show speakers of English tend to avoid using the verbs *go* and *come* after the future marker *going to*?

RQ2: More generally, how can claims of tendencies of this type be confirmed or refuted with evidence from corpora?

4. METHODOLOGY

Two corpora were selected for the basic assessment of the rule: the British National Corpus (the 'BNC': a 100-million word corpus of British English speech and writing), and the Corpus of Contemporary American English ('COCA': a 560-million word corpus of American English speech and writing). These two relatively large corpora were chosen in order to access a large sample of language, crucial when attempting to address the question of whether or not, and to what extent, a particular structure is attested in real language. Both corpora also contain both spoken and written data, which was felt desirable. The reason for consulting both the American English COCA in addition to the British English BNC was that this writer is not aware of any American English ELT publications to include the *going to* rule; the possibility of different uses in American English and British English could therefore be investigated. Both corpora were accessed through the BYU platform (www.english-corpora.org).

A number of different tests were devised in order to attempt to assess the purported rule. These can be roughly divided into three categories, as follows.

4.1. Test type 1: Raw frequency

The first type of evidence to be collected related to raw frequency. Searches were carried out in both corpora for the multi-word units (O'Keeffe, McCarthy & Carter 2007) *going to go* and *going to come*, to ascertain whether or not they were attested at all. Following this, the fifteen most common *going to* + infinitive chunks and their frequencies in the BNC and COCA were identified, in order to compare – again at the level of raw frequency – the behaviour of *go* and *come* after *going to* with the behaviour of other verbs in the same environment.

4.2. Test type 2: *go* / *come* vs. similarly frequent verbs

One problem with data related to raw frequency is that it does not address the question of a linguistic *tendency*. Information on the frequency of phenomenon X in environment Y cannot tell us alone whether that frequency is lower than might otherwise be expected: in other words, whether speakers *tend to* avoid (although do not completely avoid) X in environment Y. To address this, it was decided to consider the behaviour of two additional verbs in the same environment; the verbs *say* and *take* were selected, as they, as lemmas, have a similar overall frequency in the BNC to *go* and *come* as lemmas [for *go* and *say*,

this is 236,135 and 311,997 respectively, and for *come* and *take* it is 143,137 and 171,684 respectively (Kilgarriff 2006)].

A number of different comparisons were carried out. Firstly, the frequencies of *go* and *come*, and of *say* and *take*, were compared, in order to ascertain whether the verbs *go* and *come* were relatively less frequent after *going to* than the verb which they had been paired with. Since both *say* and *take* have a similar overall frequency to *go* and *come*, should they also have a similar frequency after *going to*, this would suggest that the 'effect' of *going to* applies equally to all four verbs, and that the rule is therefore untrue.

Again, however, there are possible objections to such a test. As Kilgarriff (2005) has argued, language is not random. Differences between the relative frequencies of *go* and *come* after *going to*, compared to those of *say* and *take*, may be influenced by other factors. In particular, given that *going to* is frequently used to talk about actions about which a decision has already been made (Carter & McCarthy 2006), it may simply be more common to talk about decisions about travel in the future (i.e. using the verb *go*) than decisions about *saying* something in the future, which would make interpretations based on the comparison described above potentially problematic. The same speculative logic could also be applied to *come* and *take*: there might be a reason why we use one more than the other – beyond the presence or otherwise of *going to* – when talking about the future.

In order to address this potential objection, a second comparison was carried out. If it is true that speakers generally prefer to talk about *going somewhere* in the future as opposed to *saying something* in the future, then this should also be reflected in the frequencies of the verbs *say* and *go* after another future marker, *will*. The relative frequencies of *will go* and *will say* were therefore established, and compared to the relative frequencies of *going to go* and *going to say*.

However, similar objections can also be raised about this kind of test. Like *going to*, the modal verb *will* is associated with particular functions, for example making offers (Carter & McCarthy 2006), and it may be the case that the frequency figures are influenced by this (we may need to communicate more frequently, for example, an offer to go somewhere as opposed to an offer to say something). The comparison in the previous paragraph was therefore extended to a further three verbs with a similar overall frequency in the BNC to *go* (according to Kilgarriff (2006), namely *get*, *make* and *take*).

4.3. Test 3: Co-occurrence information

The third type of test makes use of corpus data on co-occurrence frequency, that is to say, how frequently two elements co-occur; in the case of this study,

these two elements are a particular verb and the structure *going to*. We are essentially interested in how frequently verb form X appears in the corpus in environment Y (in other words, after *going to*), as a percentage of the total number of times that verb X appears in the corpus.

Co-occurrence information on the verbs *go* and *come* by themselves is not particularly useful and must be contextualised in some way. Two approaches were carried out to address this. Firstly, the analysis was extended to the ten most common lexical verbs in the BNC (retrieved from Leech, Rayson & Wilson 2001); both *go* and *come* are among the verbs in this list. The co-occurrence frequency of each verb after *going to*, as a percentage of overall frequency of the verb in the corpus, was calculated for all ten verbs, in order to provide a broader overall picture of the behaviour of *go* and *come* after *going to*, in comparison with the behaviour of other verbs in the same environment – essentially to ascertain whether *go* and *come* behave any differently from other common verbs when they appear after *going to*.

Following this, two statistical measures of association were employed: Point-wise Mutual Information Score and Odds Ratio (see Gries [in press] for an overview). These measures indicate whether two elements – in the present study, a particular verb, and the structure *going to* – co-occur more often or less often than expected (or at a frequency that would be expected by chance). To carry out these statistical tests, contingency tables were created for the ten verbs, considering the frequency of the verb immediately after *going to*, and its frequency in the corpus overall; following this, the statistical tests were carried out using Microsoft Excel.

5. RESULTS

5.1. Test type 1: Raw frequency

The BNC yields 584 examples of *going to go* (5.84 per million words) and 310 examples of *going to come* (3.1 per million words). Turning to American English, the number of hits in COCA for *going to go* is 13,366 (23.14 per million words), and for *going to come* it is 6,436 (11.14 per million words). Clearly, then, both forms can said to be attested – to a certain extent – in the BNC and COCA. These figures also clearly indicate that both forms are more frequent in American English than British English: *going to go* is almost four times as frequent, and *going to come* is 3.6 times as frequent. However, it should also be noted that *going to* followed by any infinitive is 2.9 times more frequent in COCA compared to the BNC: it appears 772.76 times per million words in the former, compared to 269.04 in the latter, which may go some way to explaining this difference.

As we have said, however, frequency alone cannot address the question of the 'tendency' suggested by the hedged rule; the strings *going to go* and *going to come* must be contextualised in some way in order to better judge the raw frequency figures. One way to assess the veracity of this hedged rule is to look at all *going to* + verb combinations, and see if *going to go* and *going to come* are among the most frequent of these combinations.

Table 2 shows the 15 most common *going to* + infinitive chunks and their frequencies in the BNC and COCA. As can be seen, there does not seem to be anything particularly infrequent about *going to go* and *going to come*: the string *going to go* is the eighth most frequent *going to* + infinitive chunk in the BNC, and the fifth most in COCA, while *going to come* is ranked fourteenth and eleventh respectively. Again, both forms are comparatively more frequent in American English, but neither appears to be particularly low in frequency. Furthermore, while both are much less frequent than the combinations at the top of the list, it should be borne in mind that *be* – the most common verb to follow *going to* – can be part of a passive or continuous structure with a different main verb (for example, *going to be finished*, or *going to be watching*), and that *do* and *have* – second and third on the list – can also operate as auxiliaries, meaning that alternative lexical verbs may be found (or be implicit) elsewhere in the clause.

In short, both the raw frequency data for *going to go* and *going to come*, and that same frequency data contextualised, through a comparison with other *going to* + infinitive chunks, suggest that the rule under consideration is inaccurate. Since the data from COCA show that both *going to go* and *going to come* are more frequent in American English than in British English, the rest of the analysis in this paper will consider only data from the BNC.

Table 2. The 15 most common *going to* + infinitive chunks in the BNC and COCA

| Rank | BNC | | COCA | |
|------|---------------------------|------------|---------------------------|--------------|
| | multi-word unit | frequency | multi-word unit | frequency |
| 1 | <i>going to be</i> | 5866 | <i>going to be</i> | 99962 |
| 2 | <i>going to do</i> | 1770 | <i>going to have</i> | 31037 |
| 3 | <i>going to have</i> | 1594 | <i>going to do</i> | 23478 |
| 4 | <i>going to get</i> | 1328 | <i>going to get</i> | 21119 |
| 5 | <i>going to say</i> | 606 | <i>going to go</i> | 13363 |
| 6 | <i>going to take</i> | 602 | <i>going to take</i> | 13037 |
| 7 | <i>going to make</i> | 585 | <i>going to happen</i> | 11039 |
| 8 | <i>going to go</i> | 584 | <i>going to make</i> | 9544 |
| 9 | <i>going to happen</i> | 530 | <i>going to see</i> | 7469 |
| 10 | <i>going to tell</i> | 376 | <i>going to say</i> | 7371 |

| Rank | BNC | | COCA | |
|------|-----------------------------|------------|----------------------|-------------|
| | multi-word unit | frequency | multi-word unit | frequency |
| 11 | <i>going to give</i> | 353 | <i>going to come</i> | 6427 |
| 12 | <i>going to see</i> | 347 | <i>going to give</i> | 4879 |
| 13 | <i>going to put</i> | 312 | <i>going to tell</i> | 4238 |
| 14 | <i>going to come</i> | 310 | <i>going to try</i> | 4014 |
| 15 | <i>going to ask</i> | 261 | <i>going to talk</i> | 3936 |

5.2. Test type 2: *go / come* vs. similarly frequent verbs

The second type of test sought to compare two additional verbs with a similar overall frequency in the BNC to *go* and *come*: specifically, *say* and *take*. Table 3 shows the overall frequency of all four verbs, and their frequency after *going to* in the BNC.

Table 3. Overall frequency and frequency after *going to* of *go*, *say*, *come* and *take* in the BNC

| Verb | Overall frequency | Frequency after <i>going to</i> |
|-------------|-------------------|---------------------------------|
| <i>Go</i> | 236,135 | 584 |
| <i>Say</i> | 311,997 | 606 |
| <i>Come</i> | 143,137 | 310 |
| <i>Take</i> | 171,684 | 602 |

The verbs *say* and *go* are the first and second most frequent lexical verbs in the BNC respectively. As can be seen in the table, *say* is slightly more frequent after *going to* than *go*, but the difference is small; for the rule to be true, we would expect to see a much greater difference between the two figures. We might therefore say that the frequency of *going to go* is, based on the behaviour of the verb *say* in the same environment, no lower than expected.

An analysis of *come* and *take*, however, presents a different situation. Like *go* and *say*, *come* and *take* are ranked next to one another in terms of overall frequency (*take* is the 7th most common lexical verb, and *come* the 8th most common), but the frequency of *going to come* is 310, compared to 602 for *going to know*. This is unexpectedly low: that there are only half as many examples of *going to come* as there are of *going to know* might suggest that speakers avoid using *come* after *going to* in a way that they do not do with *say*. However, the rule under examination is never stated this way; in some cases it is said to apply just to *go*, but never just to *come*. Alternative explanations take us into the realm of hypothesis: do we simply have more communicative need to talk about *taking*

rather than *coming* in the future? And might we make similar arguments, but in reverse, about *go* and *say*?

To address such hypotheses, we can also consider the modal verb *will*. Table 4 shows the raw frequencies of *go*, *say*, *get*, *come* and *take*, and their frequencies after *will* (including the contracted form 'll) and *going to*. As we have said, the aim of this comparison is to attempt to get at the question of whether *going to go* – although attested at a similar frequency to *going to say* – should actually be more frequent than it is. Such a claim would be based on the hypothesis that we are likely to have a greater, and more frequent, communicative need to talk about *going* somewhere in the future than *saying* something, and this should be reflected in the frequency data presented. Initially, the analysis suggests that this may indeed be the case: there are a total of 4,649 examples of *will go*, compared to only 1,283 for *will say*. The fact that there are fewer examples of *going to go* than *going to say* (584 vs. 606) is therefore potentially surprising. A further verb with a similar overall frequency to *go* – *get* – seems similarly anomalous: there are 4,578 examples of *will get*, compared to 4,649 for *will go*, but then over double the number of examples of *going to get* compared to *going to go* (1,329 vs. 606). The verbs *say* and *get* therefore appear to offer evidence in support of the rule: going by the data from *will* + infinitive, the string *going to go* appears to be less frequent than we might expect it to be.

The verbs *make* and *take*, however, tell a different story. There are 3,515 examples of *will make*, compared to the 4,649 for *will go*, and then 585 for *going to make*, compared to 606 for *going to go*. For *take*, the figures are 4,880 examples of *will take*, and 602 for *going to take*. These verbs therefore offer counter evidence to the rule, which would predict that there should be more examples of *going to take* and *going to make* than *going to go*, not a similar number, or fewer. Again, it is possible to speculate about the difference between *say* and *get* on the one hand, and *make* and *take* on the other. Perhaps the fact that *will* and *going to* are used to talk about the future in different ways affects which verbs appear more or less frequently after them – it can be speculated that speakers are less likely to make a prediction that somebody or someone will go somewhere than they are to express an intention with *go*.

Table 4. Raw frequency of *go*, *say*, *get*, *come* and *take*, their frequency after *will* (including its contracted form), and their frequency after *going to*

| Verb | Overall frequency | Frequency after <i>will</i> | Frequency after <i>going to</i> |
|-------------|-------------------|-----------------------------|---------------------------------|
| <i>Go</i> | 236,135 | 4649 | 584 |
| <i>Say</i> | 311,997 | 1283 | 606 |
| <i>Get</i> | 211,009 | 4578 | 1329 |
| <i>Make</i> | 208,322 | 3515 | 585 |
| <i>Take</i> | 171,684 | 4880 | 602 |

5.3. Test type 3: Co-occurrence information

We now turn to the question of co-occurrence information. Table 5 presents the ten most common lexical verbs in the BNC, showing their overall frequency, their frequency after *going to*, and their frequency after *going to* as a percentage of their overall frequency. Here the figure for 'overall frequency' is based on the lemma of each verb; in other words, the figure of 176,875 for the verb *know* includes all forms of the lemma, i.e. *know*, *knows*, *knowing*, *knew* and *known*.

Of main interest is the percentage figure in the rightmost column, and the table overall is ordered according to these figures. It is not suggested that each percentage value is meaningful in itself; the percentage figure of 0.25% for *go* cannot be said – by itself – to be a lot or a little. However, the values in relation to one another are more revealing. As can be seen, *go* and *come* fall exactly midway through the list, meaning that four of the other verbs on the list are found relatively more often, as a percentage of their overall frequency, after *going to* than *go* and *come*, and four are found relatively less often after *going to*. This again seems to suggest that there is nothing special about the behaviour of *go* and *come* after *going to*.

Table 5. Frequency of *going to* + the ten most common lexical verbs in the BNC, presented as a percentage of overall frequency of each verb

| Verb | Overall frequency | Frequency after <i>going to</i> | Frequency after <i>going to</i> as % of overall frequency |
|--------------|-------------------|---------------------------------|---|
| <i>know</i> | 176875 | 37 | 0.02 |
| <i>think</i> | 142130 | 40 | 0.03 |
| <i>say</i> | 311997 | 606 | 0.19 |
| <i>see</i> | 181549 | 347 | 0.19 |
| <i>come</i> | 143137 | 310 | 0.22 |
| <i>go</i> | 236135 | 584 | 0.25 |
| <i>make</i> | 207674 | 585 | 0.28 |
| <i>give</i> | 123533 | 353 | 0.29 |
| <i>take</i> | 171684 | 602 | 0.35 |
| <i>get</i> | 211099 | 1329 | 0.63 |

The two statistical tests – MI score and Odds Ratio – lead one to the same conclusion. Table 6 presents the MI and Odds Ratio scores for the same ten verbs. As can be seen, the MI score for both *come* and *go* is close to zero (0.466 and 0.657 respectively), suggesting a neutral relationship; the figure should be *below* zero in order to suggest that the two elements (*going to* and the verb) 'repel' one

another. The same is true for the Odds Ratio measure; the score for *go* and *come* would (in the case of this statistical measure) be below 1 if the two elements repelled one another, but in both cases it is above 1 (1.386 for *come* and 1.591 for *go*). In short, these statistical measures offer no support for the claim that speakers tend to avoid using *go* or *come* after *going to*, although they do suggest that the strings *going to know* and *going to think* are less common than might be expected.¹ It should nevertheless be acknowledged, however, that even statistical tests such as these cannot really address the question as to whether there might be other reasons (of the type discussed in Section 5.2 above) that make any particular verb more or less common after *going to*.

Table 6. MI and Odds Ratio scores for *going to* + the ten most common lexical verbs in the BNC

| Verb | (Pointwise) Mutual Information | Odds Ratio |
|--------------|--------------------------------|------------|
| <i>Know</i> | -2.906 | 0.132 |
| <i>Think</i> | -2.478 | 0.178 |
| <i>Say</i> | 0.309 | 1.245 |
| <i>See</i> | 0.286 | 1.222 |
| <i>Come</i> | 0.466 | 1.386 |
| <i>Go</i> | 0.657 | 1.591 |
| <i>Make</i> | 0.845 | 1.816 |
| <i>Give</i> | 0.866 | 1.836 |
| <i>Take</i> | 1.161 | 2.269 |
| <i>Get</i> | 2.006 | 4.194 |

6. DISCUSSION

Is it possible, based on the corpus evidence presented in the previous section, to make a definitive judgement on the *going to* rule? It may be worthwhile 'inverting' the question by carrying out a thought experiment. If a grammarian were constructing the first ever pedagogical grammar of English, and was therefore not familiar with any descriptions of English grammar (since none existed), but held a suspicion that *go* and *come* are not typically used after *going to*, would they be persuaded that this suspicion was correct by the weight of evidence presented here? It is this writer's belief that they would not be persuaded, as the

¹ As a comparison, MI and Odds Ratio scores were also obtained based on the overall frequency in the corpus of the base form of each verb (i.e. 'bare infinitive') rather than as a lemma. The results for *come* and *go* are very similar (MI for *come* = 0.377, MI for *go* = 0.328; Odds Ratio for *come* = 1.305, Odds Ratio for *go* = 1.264).

evidence supporting the suspicion is scant, and, at best, inconclusive. In short, this commonly given rule should no longer feature in explanations of the use of *going to* used by learners of EFL.

As we saw in Section 2.3, many previous studies to have used evidence from corpora to critique descriptions of grammar in ELT materials have discussed possible reasons why the authors of such pedagogical accounts may sometimes choose to ‘ignore’ empirical evidence. One commonly offered explanation is that pedagogical exigencies may at times be more important for materials designers and teachers than accuracy. Shortall (2007: 179), for example, states that “pedagogic considerations may sometimes override considerations of frequency”, arguing that there should be a role for considering both frequency data and “pedagogic necessity” when designing teaching materials. Similarly, Timmis (2015) asks whether a teacher might choose to spend more time on one area of grammar than another – regardless of which is more frequent in a corpus – because it is problematic for a particular group of learners and / or easy to illustrate in class. It is difficult, however, to be so charitable in the case of the *going to* rule. The rule appears both to be untrue and also to offer no particular pedagogical benefits – it is simply extra information for a learner to internalise, and, if it is inaccurate, this is simply wasted effort.

Why, then, did the rule ever become part of pedagogical grammar explanations of the *going to* future form, at least in British English textbooks and grammars? Figure 1 presents a Google Ngram analysis (Michel et al. 2011) for *going to go*, *going to come*, *going to take* and *going to say* over the period 1850–2008 for British English. The inclusion of the latter two strings allows us to see whether any changes in frequency in *going to go* and *going to come* are simply an indication of an overall increase in the frequency of the future marker *going to*, regardless of the verb that follows it, rather than of anything special about *going to go* and *going to come*. The Ngram suggests that there has been a historical change in usage: there is a steady increase in frequency of both *going to go* and *going to come* in both British English and American English, which becomes particularly marked in the latter half of the twentieth century. While the frequency of both *going to take* and *going to say* also increased in the same period, they both appear to have been established usages before this period.²

² One weakness of Google Ngrams is that they are based on published texts. It is not possible to be sure whether this kind of data reflects actual usage, or whether it reflects more what authors, editors and publishers felt *should* be said rather than what people actually said. If the latter is true, the increase in frequency of *going to go* and *going to say* may represent a loosening in editorial guidelines rather than a change in general usage. However, I have been unable to find any advice or instructions to avoid *go* and *come* after *going to* in any editions of influential style guides from

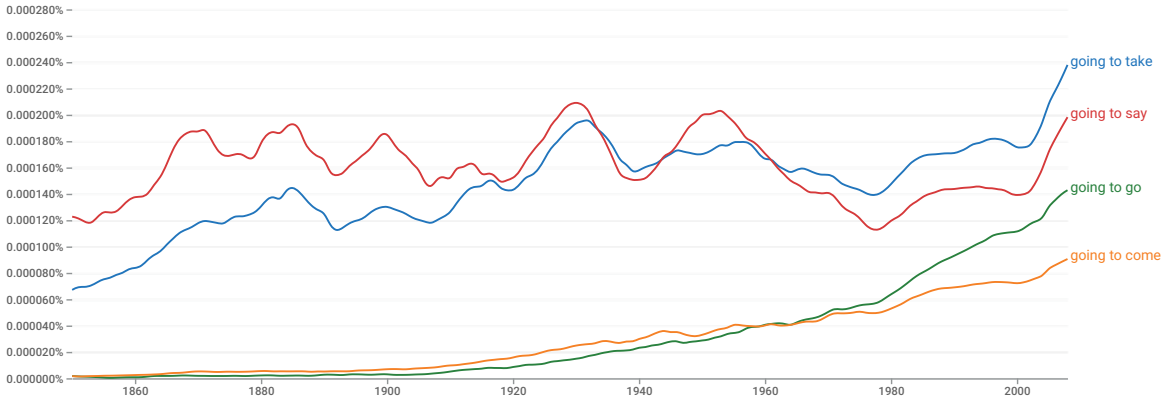


Figure 1. Google Ngram for British English of *going to go/come/take/say* over the period 1850–2008

It is therefore possible that there was once a tendency – perhaps a conscious, style-driven one – to avoid the use of *go* and *come* after *going to*, and that older pedagogical accounts such as Palmer's (1924) reflected this. However, while this tendency apparently no longer exists, ELT materials still suggest that it does. The explanation for this may have more to do with tradition than pedagogical considerations. The catalogue of grammar points typically included in ELT materials has been described as a 'canon' (Ellis 2006; O'Keeffe & Mark 2017); just as over the decades a collective agreement has developed as to which works are the most important in art and literature, it seems that within the ELT profession, an understanding has emerged as to which grammar points need to be taught, and this is well established and rarely examined. The *going to* rule is simply a part of this canon; it is repeated, uncritically, in title after title, perhaps considered true simply by virtue of the fact that it appears so frequently in print.

7. CONCLUSION

This study has attempted to shed light on the oft-repeated pedagogical grammar 'rule' that *going to* is not normally followed by *go* or *come*, assessing its accuracy with corpus evidence. The evidence examined suggests that learners should no longer be taught the rule as it does not appear to reflect actual usage. The study has also attempted to address the question of how claimed 'tenden-

the early part of the 20th century, such as Fowler's *The King's English*, the *Chicago style guide* or Gowers' *Complete plain words*.

cies' in language can be investigated, and the difficulties that can be inherent in this, suggesting different analyses that may be carried out.

The *going to* rule is indicative of what can happen once a 'canon' becomes well established: if a materials designer assumes that previously published explanations are complete and reliable, then the risk is that she or he simply repeats them uncritically. The apparent inaccuracy of the *going to* rule leads us to conclude that there could also be errors elsewhere in the canon, and that materials designers and teachers may benefit from at times taking a more critical approach to the grammatical explanations available to them, and be ready to test them against empirical evidence, such as that offered by language corpora.

As we have seen, testing claimed grammatical tendencies using corpus evidence can be complex but is not impossible. While raw frequency – even if it is low – can arguably be enough by itself to disprove outright claims that a certain structure does not occur, more sophisticated techniques may be required to test any kind of a claim of a linguistic tendency. This paper has argued that statistical tests considering co-occurrence information, and frequency-based comparisons with 'parallel structures' to that under examination (for example comparing the frequencies of *going to go* with *going to say*) can be useful in attempting to prove or disprove such claims.

REFERENCES

- Allen, W.S. (1947). *Living English structure*. Harlow: Longman.
- Azar, S. / Hagen, S. (2019). *Fundamentals of English grammar* (5th ed.). Hoboken, NJ: Pearson Education.
- Barbieri, F. / Eckhardt, S.E.B. (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech. *Language Teaching Research* 11 (3), 319–346. DOI: <https://doi.org/10.1177/1362168807077563>.
- Biber, D. / Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24 (2), 199–208. DOI: <https://doi.org/10.1017/S0272263102002048>.
- Brisard, F. (2001). Be going to: An exercise in grounding. *Journal of Linguistics*, 37 (2), 251–285. DOI: <https://doi.org/10.1017/S0022226701008866>.
- Burton, G. (2012). Corpora and coursebooks: Destined to be strangers forever? *Corpora*, 7 (1), 91–108. DOI: <https://doi.org/10.3366/cor.2012.0019>.
- Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal*, 52 (1), 43–56. DOI: <https://doi.org/10.1093/elt/52.1.43>.
- Carter, R. / McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide: spoken and written English grammar and usage*. Cambridge: Cambridge University Press.
- Close, R.A. (1970a). Problems of the Future Tense (1). *ELT Journal*, XXIV (3), 225–232. DOI: <https://doi.org/10.1093/elt/XXIV.3.225>.
- Close, R.A. (1970b). Problems of the Future Tense (2). *ELT Journal*, XXV (1), 43–49. DOI: <https://doi.org/10.1093/elt/XXV.1.43>.

- Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In: J.M. Sinclair (ed.), *How to use corpora in language teaching* (pp. 67–85). Amsterdam: J. Benjamins.
- Cook, G. (1998). The uses of reality: A reply to Ronald Carter. *ELT Journal*, 52 (1), 57–63. DOI: <https://doi.org/10.1093/elt/52.1.57>.
- Declerck, R. / Reed, S. / Cappelle, B. (2006). *The Grammar of the English tense system: A comprehensive analysis* (vol. 1). Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110199888>.
- Eastwood, J. (1994). *Oxford guide to English grammar*. Oxford: Oxford University Press.
- Eastwood, J. (2003). *Oxford practice grammar: With answers* (2nd ed.). Oxford: Oxford University Press.
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40 (1), 83–107. DOI: <https://doi.org/10.2307/40264512>.
- Fitzgerald, J. (1901). *Word and phrase: True and false use in English*. Chicago: A.C. McClurg & Co.
- Foley, M. / Hall, D. (2012a). *MyGrammarLab: Advanced C1/C2*. Harlow: Pearson Education.
- Foley, M. / Hall, D. (2012b). *MyGrammarLab: Intermediate B1/B2*. Harlow: Pearson Education.
- Frazier, S. (2003). A corpus analysis of *would*-clauses without adjacent *if*-clauses. *TESOL Quarterly*, 37 (3), 443–466. DOI: <https://doi.org/10.2307/3588399>.
- Gabrielatos, C. (2003, September 4). *Conditional sentences: ELT typology and corpus evidence*. 36th Annual BAAL Meeting. <http://eprints.lancs.ac.uk/140> [access: 23.05.2019].
- Gabrielatos, C. (2006, July 1). *Corpus-based evaluation of pedagogical materials: If-conditionals in ELT coursebooks and the BNC*. 7th Teaching and Language Corpora Conference. http://eprints.lancs.ac.uk/882/1/TALC_2006-CG.pdf [access: 23.05.2019].
- Gries, S. (in press). How to use statistics in quantitative corpus analysis. In: A. O’Keeffe / M. McCarthy (eds.), “The Routledge handbook of corpus linguistics” (2nd ed.). Oxford: Routledge.
- Huddleston, R.D. / Pullum, G.K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Jespersen, O. (1933). *Essentials of English grammar*. London: George Allen and Unwin Ltd.
- Jones, C. / Waller, D. (2011). If only it were true: The problem with the four conditionals. *ELT Journal*, 65 (1), 24–32. DOI: <https://doi.org/10.1093/elt/ccp101>.
- Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, 1 (2). DOI: <https://doi.org/10.1515/cllt.2005.1.2.263>.
- Kilgarriff, A. (2006). *BNC database and word frequency lists*. <https://www.kilgarriff.co.uk/bnc-readme.html> [access: 21.01.2018].
- Kruisinga, E. (1931). *A handbook of present-day English* (vol. II, 5th ed.). Groningen: P. Noordhoff.
- Leech, G.N. / Rayson, P. / Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus* (1st ed.). Harlow: Longman.
- Littlejohn, A. (1992). “Why are English language teaching materials the way they are?” [Unpublished PhD thesis, University of Lancaster]. www.andrewlittlejohn.net/website/books/phd.html [access: 01.11.2018].
- Locher, M.A. (2008). The rise of prescriptive grammars on English in the 18th century. In: M.A. Locher / J. Strässler (eds.), *Standards and norms in the English language* (vol. 95, pp. 127–148). Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110206982.1.127>.
- Locke, M. (1986). The Future progressive. *ELT Journal*, 40 (4), 323–325. DOI: <https://doi.org/10.1093/elt/40.4.323>.
- Michel, J.-B. / Shen, Y.K. / Aiden, A.P. / Veres, A. / Gray, M.K. / The Google Books Team: Pickett, J.P. / Hoiberg, D. / Clancy, D. / Norvig, P. / Orwant, J. / Pinker, S. / Nowak, M.A. / Aiden, E.L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331 (6014), 176–182. DOI: <https://doi.org/10.1126/science.1199644>.

- Murphy, R. (2019). *English grammar in use: A self-study reference and practice book for intermediate learners of English*. Cambridge: Cambridge University Press.
- Nesfield, J.C. (1898). *English grammar past and present*. London: Macmillan & Co., Limited.
- O'Keefe, A. / Mark, G. (2017). The English grammar profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics*, 22 (4), 457–489. DOI: <https://doi.org/10.1075/ijcl.14086.oke>.
- O'Keefe, A. / McCarthy, M. / Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Palmer, H. (1924). *A Grammar of spoken English on a strictly phonetic basis*. Cambridge: W. Heffer & Sons Ltd.
- Penston, T. (2009). *A concise grammar for English language teachers*. Wicklow: TP Publications.
- Prodromou, L. (2003). In search of the successful user of English. *Modern English Teacher*, 12 (2), 5–14.
- Quirk, R. / Leech, G. / Greenbaum, S. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
- Redston, C. / Cunningham, G. (2005). *Face2face pre-intermediate student's book*. Cambridge: Cambridge University Press.
- Redston, C. / Cunningham, G. (2016). *Face2face upper Intermediate student's book*. Cambridge: Cambridge University Press.
- Richards, J.C. / Schmidt, R.W. (eds.). (2007). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). Harlow: Longman.
- Römer, U. (2005). *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts, and didactics*. Amsterdam: J. Benjamins.
- Shortall, T. (2007). The L2 syllabus: Corpus or contrivance? *Corpora*, 2 (2), 157–185. DOI: <https://doi.org/10.3366/cor.2007.2.2.157>.
- Swan, M. (1994). Design criteria for pedagogic language rules. In: M. Bygate / A. Tonkyn / E. Williams (eds.), *Grammar and the language teacher* (pp. 44–55). New York: Prentice Hall.
- Swan, M. (2016). *Practical English usage* (4th ed.). Oxford: Oxford University Press.
- Sweet, H. (1898). *A new English grammar logical and historical* (vol. II: *Syntax*). Oxford: Oxford University Press.
- Thomson, A.J. / Martinet, A.V. (1960/1969). *A practical English grammar*. Oxford: Oxford University Press.
- Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. London: Routledge, Taylor & Francis Group.
- Tregidgo, P.S. (1980). 'Shall', 'will', and the Future. *ELT Journal*, 34 (4), 265–269. DOI: <https://doi.org/10.1093/elt/34.4.265>.
- Vandenhoeck, T. (2018). The Past perfect in corpora and EFL / ESL materials. *Research in Language*, 16 (1), 113–133. DOI: <https://doi.org/10.2478/rela-2018-0005>.

Received: 04.08.2020; revised: 15.03.2021

GRAHAM BURTON

Facoltà di Scienze della Formazione, Libera Università di Bolzano

grahamfrancis.burton@unibz.it

ORCID: 0000-0002-8501-9207